



Theses and Dissertations

2021-04-11

The Effects of Dynamic Written Corrective Feedback on the Accuracy and Complexity of Writing Produced by L2 Graduate Students

Lisa Rohm
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Arts and Humanities Commons](#)

BYU ScholarsArchive Citation

Rohm, Lisa, "The Effects of Dynamic Written Corrective Feedback on the Accuracy and Complexity of Writing Produced by L2 Graduate Students" (2021). *Theses and Dissertations*. 9408.
<https://scholarsarchive.byu.edu/etd/9408>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

The Effects of Dynamic Written Corrective Feedback on the Accuracy
and Complexity of Writing Produced by L2 Graduate Students

Lisa Rohm

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Grant Eckstein, Chair
Norman W. Evans
K. James Hartshorn

Department of Linguistics
Brigham Young University

Copyright © 2021 Lisa Rohm

All Rights Reserved

ABSTRACT

The Effects of Dynamic Written Corrective Feedback on the Accuracy and Complexity of Writing Produced by L2 Graduate Students

Lisa Rohm
Department of Linguistics, BYU
Master of Arts

What started as a discussion of the efficacy of explicit grammar instruction has over time led to a debate about the need for and effectiveness of written corrective feedback (WCF) within the field of English language teaching (ELT). Dynamic written corrective feedback (DWCF) is a relatively new strategy developed by Dr. Norman Evans to provide WCF to English as a second language (ESL) students through systematic, coded feedback. While previous studies on DWCF have looked at its effects among other groups, few studies have examined DWCF in the context of ESL students studying at the graduate level. This study analyzes the linguistic accuracy and lexical and syntactic complexity of these students before and after a fourteen-week DWCF intervention.

Keywords: DWCF, written corrective feedback, graduate students, writing, grammar, complexity

ACKNOWLEDGEMENTS

My thanks to all who have helped me throughout my master's program and work on this thesis. To the student participants, thank you for your time and willingness to be part of this research. Special thanks to my thesis chair, Grant Eckstein, for his expertise and guidance and to my committee members, Norman Evans and James Hartshorn, who pioneered DWCF and provided valuable feedback. To my parents, thank you for your support throughout my life, including in my education. Thank you as well to my dear friend, Alyssa Alexander, and my beloved husband, Trenten Rohm, both of whom have supported and encouraged me throughout my studies.

CONTENTS

TITLE	i
ABSTRACT	ii
ACKNOWLEDEMENTS	iii
CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIURES	vii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW.....	4
Should educators give written grammar feedback?.....	4
How should educators give grammar feedback?	7
Previous studies on DWCF	9
This study	16
Research Questions.....	17
CHAPTER 3: METHODS.....	19
Participants.....	19
Writing Course	21
Materials.....	22
Procedures	22
Data Analysis.....	25

CHAPTER 4: RESULTS.....	30
Research Question 1: Accuracy	30
Research Question 2: Complexity	32
Research Question 3: Fluency.....	37
CHAPTER 5: DISCUSSION AND CONCLUSION.....	40
Discussion.....	40
Research Question 1	40
Research Question 2	41
Research Question 3	43
Discussion Summary.....	44
Delimitation and Suggestions for Future Research.....	45
Pedagogical Implications	48
Conclusion.....	49
References	50
Appendix A List of Prompts	57
Appendix B Treatment roup Materials	58
Appendix C Control roup Instructions	61

LIST OF TABLES

Table 1	<i>Student Demographics</i>	20
Table 2	<i>Variables</i>	26
Table 3	<i>Accuracy Results</i>	30
Table 4	<i>Syntactic Complexity Results</i>	33
Table 5	<i>Lexical Complexity Results</i>	35

LIST OF FIGURES

Figure 1 <i>DWCF Process</i>	23
Figure 2 <i>Global Errors</i>	31
Figure 3 <i>Local Errors</i>	31
Figure 4 <i>Errors Mechanical</i>	32
Figure 5 Mean Length of Sentence	33
Figure 6 Mean Length of T-unit.....	34
Figure 7 Complex Nominals per Clause.....	34
Figure 9 <i>Lexical Density</i>	35
Figure 10 <i>Lexical Sophistication</i>	36
Figure 11 <i>Type-token Ratio</i>	36
Figure 12 <i>Word Count</i>	38
Figure 13 <i>Sentence Count</i>	38
Figure 14 <i>T-unit Count</i>	39
Figure A1 <i>Student Paragraph Examples</i>	58
Figure A2 <i>Key to Codes</i>	59
Figure A3 <i>Example Error Log</i>	60
Figure C1 <i>Paragraph Instructions Part 1</i>	61
Figure C2 <i>Paragraph Instructions Part 2</i>	61

The Effects of Dynamic Written Corrective Feedback on the Accuracy
and Complexity of Writing Produced by L2 Graduate Students

CHAPTER 1: INTRODUCTION

Grammar is an essential aspect of language learning and of high concern to second language (L2) learners. For many English language learners (ELLs), proper grammar, together with good writing skills, allows them to be understood, successful, and respected by their peers in academia and the workforce. In this study, the effects of a methodological approach to written grammar feedback known as dynamic written corrective feedback (DWCF) were examined using L2 graduate (master's and Ph.D.) students. This study looked not only at the grammatical effects of this approach, but also its syntactic, lexical, and fluency effects on student writing since these too are essential to ELLs' success.

In order to do research related to grammar, it must first be defined. It has been noted that grammar may mean different things for native and nonnative speakers, with the former considering it “issues with punctuation, conjunctions, prepositions, confused words, and agreement issues” while the latter “[includes] verb tense, verb form, sentence structure, word order, and article concerns” (Eckstein, 2018a, see also Connors & Lunsford, 1988; Gillespie & Lerner, 2000; and Lane and Lange, 2011). For the purpose of this study, both of these definitions apply, allowing for a larger number of errors to be addressed by DWCF than by either alone. Simply put, this study addresses all errors that are not related to the overall organization or structure of an essay or paragraph under the umbrella of grammatical errors.

In addition to looking at grammar, this study will also explore the effects of DWCF on other aspects of writing as, over time, it has begun to be examined as not just a tool for accuracy

improvements but as a tool for potentially improving all aspects of writing. While the cognition hypothesis (Robinson 2001, 2003, 2005, 2007a, 2007b) suggests that this is both possible and likely, the trade-off hypothesis (Skehan, 1998) argues against the likelihood of students showing universal improvement over the same period of time. Thus, with two opposite rationales for potential effects on other aspects of writing and mixed research results thus far (Hartshorn & Evans, 2015; Eckstein et al., 2020), this study aimed to continue investigations into DWCF's effects on syntactic and linguistic accuracy.

DWCF has been researched using pre-matriculated IEP, undergraduate, and graduate students. Results thus far (from research both with and without control groups) indicate positive effects on grammatical accuracy at the IEP and undergraduate levels (Hartshorn et al., 2010; Evans et al., 2010; Lee, 2009; Evans et al., 2011; Hartshorn & Evans, 2015). Research with graduate students, however, has yielded no significant accuracy results, suggesting a possible ceiling effect (Kurzer, 2018 Eckstein et al., 2020). Moreover, it is unclear how DWCF affects different groupings of error types (Hartshorn & Evans, 2012, Kurzer, 2018, Eckstein et al., 2020). Studies have also looked at complexity and fluency, with results indicating either that DWCF has no statistically significant effect or may positively affect some areas with others showing a decrease in complexity scores (Hartshorn & Evans, 2015; Eckstein et al., 2020). Further research, including this study, look to expand the current understanding of DWCF's effects, especially among high-level students, and to investigate the potential ceiling effect and limitations of DWCF's usefulness regarding accuracy.

Although L2 graduate students have been the focus of a single DWCF study (Eckstein et al., 2020) in the past, that study did not include a control group and addressed a different set of research questions than the present study. Studying L2 graduate students may help researchers:

a) learn more about DWCF and what proficiency levels it is useful in (i.e., if there truly is a ceiling effect) and b) address student needs. Knowing when DWCF is shows accuracy gains allows educators to make informed decisions about its usage. Additionally, the second point is supported by research indicating that L2 graduate students—many of whom are working towards publishing in their field (Kuzhabekova, 2020)—still lack the writing and grammatical skills they need (Bitchener & Bastrukmen, 2006; Hyland & Milton, 1997; Kuzhabekova, 2020; Leki, 2010; Ma, 2019; Powers & Nelson, 1995). In order to know if DWCF is effective in addressing these student needs, this research looks at several potential effects on student complexity and fluency.

CHAPTER 2: LITERATURE REVIEW

Should educators give written grammar feedback?

In any discussion of DWCF, it is necessary to start with The Grammar Correction Debate (1996-2004). This refers to a series of four articles which discussed whether grammar correction, broadly meaning any form of correcting students' grammatical errors, is helpful or harmful for L2 student writers. The debate began with an article by Truscott (1996). Through reviews of prior research on written corrective feedback and his own logical reasoning, Truscott came to the conclusion that grammar correction was one of either two things: useless, but neutral, or harmful. He concluded that because there was a lack of convincing empirical evidence in the then-current research, "grammar correction has no place in writing classes and should be abandoned" (p. 361) until further research had shown both if it is useful and what parameters are necessary for it to be useful in.

In response to Truscott's opinion on error correction, Ferris published an article (1999) in which she countered several of Truscott's claims. She addressed both the fact that he did not differentiate between the various methods of "grammar correction" (she herself added the quotation marks to indicate the ambiguity of this term), despite the fact that surely some approaches are more useful than others, and criticized his review of the current literature, pointing out studies which would indicate the usefulness of grammar correction. She further argued that discontinuing this sort of feedback could have a negative effect on the motivation of expectant students. Regardless, she did agree that more research was needed and that the current literature was lacking.

This debate wrapped up with two final articles from Truscott and Ferris respectively. In the final article of this debate, Ferris (2004) noted that "two points on which [Truscott] and [her

agree] are (a) that the research on error correction in L2 writing is indeed insufficient and (b) that the ‘burden of proof’ is on those who would argue *in favor* of error correction” (p. 50, original emphasis). However, once again they differed in their view of the future. Truscott’s articles were heavily influenced by Noam Chomsky’s ideas of universal grammar (1957) and the idea that truly learning a language is different than temporarily being able to produce the correct forms (what Truscott refers to as *learning* and *pseudolearning* respectively) (1996). His views were therefore heavily influenced by first language (L1) acquisition theory and he still maintained his opposition to error correction.

On the other hand, Ferris operated under the idea that being able to self-correct is part of a process that leads towards long-term improvement, i.e., true learning. She used research from the field of second language acquisition (SLA) to support her ideas and theoretical underpinnings. This SLA research focused on the connection between the ability to self-correct and long-term improvement. She noted that “Recent second language acquisition (SLA) research on Focus on Form (in both written and spoken language) strongly suggests that adult second language acquirers in particular need their errors made salient and explicit to them so that they can avoid fossilization and continue developing linguistic competence” (p. 54, see also Doughty & Varela, 1998; Doughty & Williams, 1998; Ellis, 1998; James, 1998; Lightbown, 1998; Lyster & Ranta, 1997; Tomasello & Herron, 1989). Ferris’ article also countered the affectivity argument against error correction—which claims that students may be negatively impacted by error correction (see also Ferris, 2009)—by drawing attention to student desire for correction: “However, from an affective standpoint, students’ strongly held opinions about this issue may influence their success or lack thereof in the L2 writing class. Thus, the existing research on

student views predicts that the presence of error feedback may be beneficial and its absence may be harmful” (Ferris, 2004, p. 55).

In the end, Truscott’s Chomskian views of language acquisition left him unconvinced of grammar correction’s effectiveness and pessimistic about future outcomes while Ferris’ SLA-informed views resulted in her encouraging further research and noting “positive indicators that error feedback may not only be helpful but necessary” (p. 56). Thus, while their analyses of the current research and outlooks towards the future differed, the Grammar Correction Debate ended with both researchers making a plea for further research and set the stage for more research to come.

After the Grammar Correction Debate, WCF was researched with some mixed results. While some research still argued against it (see Truscott & Hsu, 2008), most research yielded positive results in favor of WCF. A meta-analysis of 21 studies found “[indications] that written corrective feedback does have a substantive effect on L2 written accuracy” (Kang & Han, 2015). Eventually, the growing evidence caused the debate over whether WCF should be done to instead focus on the most effective methods of WCF (see also Hartshorn et al., 2010; Evans et al., 2010).

Further research into Skill Acquisition Theory also provides support for Ferris’ SLA-based position that linguistic knowledge is needed for ELLs. Skill Acquisition Theory posits that any skill is learned first as declarative knowledge (knowledge of or about a skill); then applied as procedural knowledge (using the skill); and then, through large amounts of practice, developed into automaticity, i.e., being able to do the task almost without thinking (DeKeyser, 2007). Particularly of interest in this theory “...is that this whole sequence of proceduralization and automatization cannot get started if the right conditions for proceduralization are not present (the

declarative knowledge required by the task at hand and a task setup that allows for use of that declarative knowledge)” (DeKeyser, 2007, p. 98). Supporting this is additional research which found “that skill acquisition involves development of a complex set of strategies based on use of rules and retrieval of examples” (Anderson, Fincham, & Douglass, 1997). Thus, skill acquisition theory provides a strong rationale for correcting student errors and teaching them to self-correct as part of the declarative and procedural steps in the skill acquisition process.

How should educators give grammar feedback?

Dynamic Written Corrective Feedback (DWCF) was first introduced as a feedback method in 2010 (Hartshorn et al.) in response to the researcher beliefs that WCF was not a question of “*whether* it was beneficial, but rather *how* to use it effectively to help their students write more accurately” (p. 85, original emphasis; see also Evans et al., 2010). These researchers asserted that proper "instructional methodology" (p. 85) would contribute to the success of WCF. By examining both the current research on if and how WCF benefitted students and the known difficulties of implementing WCF, they created DWCF intending to provide language learners with effective feedback while mitigating the negative effects (being overwhelming, time-consuming, etc.) on students and teachers. Additionally, this instructional strategy was designed to improve *student* linguistic accuracy, not just the accuracy of the single piece of writing receiving feedback.

DWCF uses indirect feedback “provided in the form of coded symbols that identify the error type and its location” (Hartshorn et al., p. 88, 2010) to mark all student grammatical errors in a piece of writing and then asks them to revise and resubmit their writing with the student choosing how to correct the highlighted errors. Research and theory into WCF favor indirect, explicit (coded) feedback for L2 learners (see Kurzer, 2018), providing a pedagogical foundation

for this method. After students receive feedback and revise, the cycle is then repeated as many times as needed to reach adequate accuracy.

As a method, DWCF was intended to address teacher needs by being learnable (for both teachers and students) without extensive training; easily integrating into preexisting courses; and having research indicating significant effects on accuracy at the time it was introduced.

Additionally, it was designed to tackle many of the logistical difficulties of giving WCF, by being "manageable, meaningful, timely, and constant" (Hartshorn et al., 2010, p. 451; see also Evans et al., 2010) while maintaining a theoretical basis. These four principles have become an essential part of DWCF and are defined in particular ways relative to it.

In the context of DWCF, manageable means that both the student and teacher are able to complete the writing and feedback without being overwhelmed. DWCF aims to do this by shortening the writing task to 10-minute paragraphs. This is meant to maintain the authenticity of the feedback by providing "unfocused feedback"—feedback that addresses all error types—as opposed to "focused feedback," which limits the feedback to only a few chosen error types as is sometimes favored in WCF (see Hartshorn & Evans, 2012). It is also designed to give teachers and students sufficient time to give, understand, and utilize feedback.

Being meaningful means that students are able to understand the feedback and how to use it and then do so. As part of this, students are expected to "invest in the learning process by reasoning through their errors" (Evans et al., 2010, see also Ferris, 2006). It is the students that choose how to correct their errors; instructors only indicate the type of error through a series of codes that students have been taught to recognize and understand. Students are also required to keep track of their errors, allowing them to see which types of errors are most prominent in their writing. In its original form, DWCF also required instructors to provide an overall writing score

so that students would receive feedback on more than just what errors were made (see Hartshorn et al., 2010).

The timely aspect of DWCF refers to using the shortest possible amount of time between students completing a task and receiving feedback, intended for fast, frequent feedback. In theory, this maximizes the amount of feedback that students receive while in a course or program that utilizes DWCF. The original study (Hartshorn et al., 2010) defined timely as being the next class period.

Constant is the final core tenant of DWCF. Feedback is considered constant when students are writing and receiving feedback close to every class period, providing them with a steady stream of feedback. This consistency is meant to help students identify what they are struggling with and strive to improve it. It also intended to prevent students from going long periods without receiving feedback and allow for consistent practice.

Previous studies on DWCF

As a relatively new form of WCF, the research on DWCF is still limited. The first published article (Hartshorn et al., 2010), introduced DWCF as an instructional strategy and looked at the effects of DWCF on pre-matriculated students of varying proficiency levels. All 47 students (28 treatment, 19 control) were enrolled at an intensive English program (IEP). After 15 weeks of daily (Monday-Thursday) DWCF intervention, the treatment group showed significant accuracy gains while the control group showed decreases in accuracy (for this study, accuracy was measured by the number of error-free T-units divided by the total number of T-units). The study also looked at rhetorical competence using a modified TOEFL iBT rubric, writing fluency using word count, and writing complexity using the average length of T-units, but found no significant differences between groups. Thus, this study provided evidence that the grammatical

accuracy of IEP students from a range of proficiency levels was improved by 15-weeks of DWCF, but no evidence of its influence on other aspects of writing.

Later that same year, another study was published by Evans et al., 2010 who emphasized that WCF research should look into specific factors (what they referred to as learner variables, situational variables, and methodological variables) to find what types of feedback works for specific groups of students (such as L2 versus L1 students). For this study, IEP students were again used, but only at the advanced-low level. After both groups concluded 13-weeks of DWCF, student accuracy (measured by error-free clauses out of total number of clauses) improved significantly. While this study also included holistic scores, no rubric reliability was established, the ratings were not done blind, only one rater looked at each essay, and the researchers themselves emphasized the subjectiveness of those scores. In the end, this study provided evidence that grammatical accuracy can improve while doing DWCF but lacked a control group to help indicate that the improvement was due to DWCF and not another factor such as time.

The next publication on DWCF (Evans et al., 2011) was a replication of the original 2010 DWCF study, with the major difference being that these students were L2 university undergraduate students instead of pre-matriculated IEP students. Using the same time period of 15 weeks and the same measure of accuracy (error-free T-units to total T-units), this study once again found that the 16 treatment group students improved in overall accuracy while the 14 control group students had decreased scores. Between the two groups, the treatment group had significantly better scores than the control group, but this study once again found little evidence for effects on student fluency and complexity. This would be the first DWCF study to provide evidence of DWCF's effectiveness on grammatical accuracy at the undergraduate level.

After multiple studies had provided evidence of DWCF's potential for improving overall accuracy, studies began to also look more closely at its effects on specific error families, writing fluency, and lexical and syntactic fluency. The first of these (Hartshorn & Evans, 2012), sorted errors into "error families" (p. 227) in order to see the degree of effectiveness that DWCF had on each. These three categories—grammatical, local, and mechanical errors—were further divided into individual error types: "The grammatical error family included sentence structure errors, determiner errors (e.g. articles, possessive nouns and pronouns, numbers, indefinite pronouns, and demonstrative pronouns), verb errors (e.g. subject-verb agreement, verb tense, and other verb form problems), numeric shift errors (e.g. count-non-count, singular-plural), and semantic errors (e.g. awkwardness, insertion / omission, unclear meaning, and word order). The lexical error family included word choice errors, word form errors, and preposition errors. The mechanical error family included errors in capitalization, indentation, non-sentence level punctuation, and spelling" (p. 227). These were examined over a 15-week period using IEP students (19 treatment students and 28 control group students) at the IEP's uppermost proficiency level. Accuracy for this study was again based on error-free T-units. All error families showed overall improvement for the treatment group, although there were varying levels of effectiveness on specific error types. The article concluded that "First, results have shown that a systematic application of the principles behind skill acquisition theory may have a positive effect on the accuracy of L2 writing for both non-grammatical and grammatical errors without undermining rhetorical competence. Second, the results underscore the assertion that focused WCF may not be the only appropriate form of feedback for every learning context: practitioners should be encouraged to explore what may be best for their specific learners" (p. 239). In the end, this study found support for the theoretical principles behind DWCF and—as a first look

into specific error types—provided evidence that DWCF is more effective for particular errors, while still benefitting overall accuracy.

After the initial wave of positive results, DWCF began to be studied using students who were learning other languages than English. Akiyama and Fleshler's 2013 research used beginning Japanese as a foreign language (JFL) students. Their study differed from other DWCF in that both groups received coded feedback, but the control group was not required to figure out the correct answer by themselves or submit multiple drafts. In addition to looking at the grammatical effects of DWCF (specifically on particles and predicates), they examined how overall writing quality was affected and student opinions on DWCF. First, the accuracy results (based on error-free T-units to errored T-units) further reinforced the notion that DWCF has differing effects on specific error types. While particle accuracy significantly improved on grammar exercises through DWCF, it had little effect on the study's essay tests, which were spontaneous writing. Conversely, predicates significantly improved for essays and—although they did not test predicates on grammar exercises so statistical significance could not be determined—there was an observable benefit to the treatment group. These findings suggest that DWCF benefits students' explicit knowledge of grammar and that DWCF exercises can, in some cases, improve spontaneous writing outside of DWCF exercises. Next, they found little difference in overall writing quality between the two groups. In the end, their research concurred with finding from Hartshorn et al. (2010) regarding accuracy improvement versus overall improvement.

Other studies brought different perspectives about DWCF to light. In an unpublished master's thesis, which replicated much of an earlier study done by Hartshorn (2008), Lee (2009) looked at 53 students (35 treatment, 18 control) from five different English proficiency levels.

Like Evans et al., 2010, Lee used the ratio of error-free clauses to errored clauses as her accuracy measurement. Noting that her advanced treatment students surpassed their control group (whose accuracy decreased) while her intermediate-high students in both groups improved, Lee suggested that “more proficient students can benefit more from DWCF” (p. 64) while still concluding that it offered benefits to student accuracy in general.

The first longitudinal DWCF study was published in 2015 (Hartshorn & Evans) using a 30-week period to examine DWCF’s effects on accuracy (as measured using error-free clause ratios) as well as other aspects of writing. This included “rhetorical appropriateness,” measured using Hartshorn et al. (2010) ’s adjusted iBT TOEFL rubric; fluency, measured through overall word count; two measures of complexity: mean length of T-unit (MLTU) and clause to T-unit ratio; and lexical development, looking at a variety of lexical measures such as usage of frequent words, type-token ratio, and lexical density. As seen in previous studies, while the treatment group showed improved accuracy compared to their peers, other variables showed little to no difference between the two groups.

The largest and most comprehensive DWCF study (Kurzer, 2018) once again examined error families split into global, local, and mechanical errors by looking at 325 ESL writing class students of varying proficiency levels over three terms. Results were examined through several lenses. First, Kurzer looked at the likelihood that DWCF improved student self-editing abilities. He did so in two ways: by looking at the number of their diagnostic paragraph errors students could correct at the end of the quarter, resulting in significant gains for the treatment group, and by examining the treatment group’s first and second paragraph drafts, with results “[suggesting] that students in each of the levels (a) responded differently to the treatment and (b) were able to self-edit their individual paragraphs differently based on those levels, which is to be expected

given the different language proficiencies seen. The improvements between drafts also suggest that students who experience DWCF are capable of effectively self-editing, reinforcing the conclusion that DWCF may contribute to learners' autonomy as accurate writers" (p. 19).

Second, this research looked at overall accuracy improvement with the treatment group showing both statistical significance and large effect sizes. Finally, the research examined effects on specific error families sorted into global, local, and mechanical errors like Hartshorn and Evans (2012) but with minor differences. Kurzer stated that "The investigation of the errors in terms of global, local, and mechanical for this study helps determine the impact DWCF may have on the student participants' comprehensibility" (p. 12, see also Bates et al., 1993; Bitchener & Ferris, 2012; Burt & Kiparsky, 1972; Hendrickson, 1980) with global errors affecting meaning, local errors remaining understandable despite being grammatically incorrect, and mechanical errors reflecting mostly on spelling and punctuation (see also Bates et al., 1993; Lane & Lange, 2012).

In this study, students showed improvement across all error categories. There are several aspects to consider when looking at these results, particularly that accuracy was measured differently in this study than in all previously mentioned studies. Instead of error-free T-unit ratios, this study used errors per 100 words. Kurzer suggests that error-free T-unit ratios "may conflate accuracy and fluency, or at least present only one possible aspect of accuracy" (see also Larsen-Freeman, 2009). He also notes that looking at error-free T-units often includes not knowing the exact quantity of each specific error type within the broader families of global, local, and mechanical errors (see also Polio & Shea, 2014). Other aspects to consider include that the treatment groups for each proficiency level differed slightly (such as with the advanced students receiving fewer DWCF cycles overall) and that students were enrolled in writing courses. While these other factors may also have contributed to student improvement, variations between the treatment and

control groups indicate that DWCF likely contributes to better self-correction and increased accuracy for all error families.

The most recent DWCF study (Eckstein et al., 2020) differed from previous studies by being the first to look at graduate-level students and examining not just grammatical accuracy but also lexical and syntactic complexity. This study also explored a key tenet of DWCF by looking at two treatment groups: those who received feedback in a “timely” manner (within a week of writing) compared to those who received delayed feedback. While there were small differences between the groups, there was no significant difference between each groups’ accuracy as measured by total (normalized) error counts (similar to Kurzer, 2018). Additionally, while both groups decreased somewhat in overall lexical and syntactic complexity, the timely group was found to have an average sentence length of approximately double that of the timely group and the timely group wrote twice as many complex nominal clauses and had a higher overall word count. While the study provided some insight, more research is needed to determine the full effects of DWCF on lexical and syntactic complexity and the study’s lack of a control group raises the question of whether DWCF in general has a significant effect on graduate students’ accuracy.

When looking at all of these studies together, there is evidence that DWCF is beneficial to the general written accuracy of pre-matriculated IEP students (Hartshorn et al., 2010, Evans et al., 2010, Lee, 2009) and undergraduate university students (Evans et al., 2011, Hartshorn & Evans, 2015), but no conclusive evidence whether it is able to help graduate-level students. Additionally, there is mixed results on the effects of DWCF on specific accuracy categories (Hartshorn & Evans, 2012, Kurzer, 2018, Eckstein et al., 2020). Finally, the current evidence regarding DWCF’s effects on complexity and fluency aspects of writing is limited and

inconclusive as some studies have seen an improvement to some measures of complexity and fluency while others showed no change or a decrease in other measures (Hartshorn & Evans, 2015; Eckstein et al., 2020). These studies provide a foundation showing that DWCF has its place in language teaching while also leaving room for further studies to investigate the full extent and boundaries of its effectiveness.

This study

Despite years of research of WCF, many teachers are still uncertain about sound pedagogical approaches for giving students WCF and research is only beginning to explore whether DWCF is an effective option. Continued testing is needed to probe the parameters of DWCF's effectiveness and usability and to fully understand its effects on complexity and fluency, not just grammatical accuracy. This present study is designed to help fill in the gaps in the existing research. First, the research looks at DWCF's effects on accuracy for graduate-level students. The motivations, content knowledge, and proficiency level of graduate-level writers differentiate them from the other learners who have participated in the majority of DWCF research thus far. Moreover, the only research using these students did not include a control group of students who did not receive DWCF. This study aims to provide insight into how DWCF changes graduate student writing by comparing a treatment group of students who received DWCF feedback to a control group with no exposure to it.

Additionally, this study continues the research of Eckstein et al. (2020) to further investigate the effects of DWCF on linguistic and syntactic complexity in order to give a more complete view of how DWCF influences student writing. Only a few studies have looked at the other potential benefits and drawbacks that may come from using DWCF. This interplay is important to investigate due to its initial suggestion going as far back as the argument by

Truscott (1999) against correction and the existence of theoretical models which support multiple possibilities.

There are two competing theories regarding student improvement. First, it has been noted that “to improve in one area often seems to be at the expense of improvement elsewhere” (Skehan, p. 112, 1998). Essentially, some hypothesize that it is unlikely for students to see universal improvements occurring at the same time and that teachers and students must “trade-off” between progress in one area with regression or stagnation in another. In particular, complexity, fluency, and accuracy have been noted as being impacted by this effect (see Skehan, 1998). This trade-off hypothesis is in contrast to the cognition hypothesis (Robinson 2001, 2003, 2005, 2007a, 2007b), which argues that students can improve both their accuracy and complexity when given increasingly cognitively heavy demands to match more complex tasks (see Robinson, 2007). Depending on which of these theories is supported by DWCF, educators may need to choose which aspects of writing their students need the most help with in order to decide whether to use DWCF.

Essentially, what is known thus far about DWCF’s effects on writing complexity and fluency are limited to three studies (Hartshorn et al., 2010; Hartshorn & Evans, 2015; Eckstein et al., 2020), all of which appear to support the trade-off hypothesis due to a lack of improvement or mixed results, with some complexity aspects improving and others worsening. By providing additional research on who benefits and how DWCF influences writing complexity and fluency, this research aims to help further establish the parameters of DWCF’s effectiveness and usefulness. This in turn will empower teachers to make better-informed choices about whether DWCF is suitable for their particular group of students.

Research Questions

1. What effect does a 14-week DWCF intervention have on L2 graduate level (master's and PhD.) students' linguistic accuracy as viewed through the lens of three error families: global, local, and mechanical errors?

Based on past research results, we expect to see a reduction in the overall number of grammatical errors in the treatment group. However, as past research is somewhat inconclusive, we expect some error families to benefit more from DWCF than others. We are also uncertain that any changes will be significant with this group due to their high proficiency level and position on the learning curve.

2. How do lexical and syntactic complexity change over the course of 14 weeks of DWCF?
Here we do not expect to see statistically significant differences between students before and after the intervention as there has been little evidence that DWCF affects these aspects of writing.

3. How is writing fluency affected by DWCF?

Previous studies have not yet shown any effects on written fluency, so we do not expect to see many changes in this area. If there are significant effects, we predict a decrease in student fluency due to increased attention to grammatical correctness.

CHAPTER 3: METHODS

This research followed a quasi-experimental design with nonrandom treatment and control groups being measured and compared to each other. This study compared a treatment group of international master's and Ph.D. students who received approximately 14 weeks of DWCF intervention (a semester's worth) as part of a graduate level writing course to international graduate students who did not complete this course.

Participants

The treatment group comprised of 23 international students ages 24 to 39 years old. All were enrolled in graduate school at Brigham Young University (12 master's students and 11 doctoral students) at the time of this study with 11 language backgrounds represented. These students were chosen using convenience sampling of three semesters of an intact linguistics course for international graduate students. They are largely characterized by being highly motivated to learn and improve. Several of them had already published academic papers in English and/or other languages, and all were working towards theses, projects, or dissertations.

The control group was a group of 9 volunteers, also international graduate students, who agreed to complete the same type of pre and posttests that the treatment group completed as part of their coursework. These were completed about 15 and a half weeks apart from each other. These students share many of the same traits as the treatment group but were instead sampled from the departments on campus through outreach efforts inviting international students to participate in a research study in return for writing feedback and modest compensation. Like the control group, these students were enrolled at Brigham Young University and met the same admittance and language skill criteria. These students did not take the same writing class, but like the treatment group, they did take other graduate-level classes at university.

Table 1

<i>Student Demographics</i>		
	Treatment Group	Control Group
Number	23	9
Sex	9 female, 14 male	3 female, 6 male
Native Language	Chinese, Portuguese, Haitian Creole, Korean, Hindi, Bengali, Spanish, Turkish, Korean, Telugu, Persian	Chinese, Urdu, Farsi/Persian, Spanish, Portuguese, Sinhala, Korean
Program	Second Language Teaching, Public Administration, Public Health, Mass Communications, Civil Engineering, Chemistry, Business Administration, Physiology and Developmental Biology, Media and Performance Studies, TESOL, Physical Medicine and Rehabilitation, Physics, Exercise Sciences, Wildlife and Wildlands Conservation,	Chemical Engineering, Social Work, Organic Chemistry, MBA, Accounting, Analytical Chemistry
Degree	12 master, 11 doctoral	4 master, 4 doctoral, 1 unreported

Group demographics. Number, sex, native language, program, and degree.

It should be noted that although education level and proficiency are not the same, student entrance into graduate programs does reflect on their proficiency. In order to be admitted into their graduate programs, these students all received a score that met the university admissions requirement on an English proficiency tests (such as the TOEFL or IELTS exams). BYU Graduate Studies (2021b) states that all graduate students are required to have at least an overall score of 85 on the TOEFL iBT, with sub scores of at least 21-22 depending on the test section, or comparable scores on other English proficiency tests. Most departments also require students to take the GRE and/or GMAT tests (BYU Graduate Studies, 2021a). While different programs place different emphases on the various sections of the GRE or GMAT, these students' ability to receive admissible scores on the GRE or GMAT in English and at least the minimum required

scores on an English proficiency test suggests a higher proficiency level than has been looked at with undergraduate and IEP students. The hard work, dedication, and high motivation required to enter these programs are also commensurate of effective language users and suggests a high proficiency level.

Writing Course

Students in the treatment group were part of a writing course designed to help international graduate students improve their writing. The course focused largely on teaching students how to properly write a literature review with many of the semester's assignments focusing on the different types of paragraphs and culminating in a full literature review that students submitted at the end of the semester. Additional assignments and class activities addressed other needs specific to international graduate students such as improving grammatical accuracy and presenting their research. Class topics were flexible—though mostly focused on research and publication—and students were encouraged to ask about any topic. Each semester also included a handful of grammar lessons based on student need (as shown in their DWCF paragraphs). Overall, the course placed the most emphasis on writing for research and publication with secondary emphasis on grammar and other topics related to graduate life and publication.

In every semester, this course was taught by the same linguistics professor and teacher's assistant (TA). In this way, there was a reasonable amount of consistency between classes, with some variation as would be expected in different semesters of a course. The instructor possessed a substantial amount of experience working with and researching L2 writers and was already familiar with the current DWCF literature, having published on it before. The TA was a graduate student studying TESOL with prior experience using DWCF feedback in the L2 classroom at the

university's IEP.

Materials

The treatment group was asked to write for 10-minutes twice a week in response to 22 prompts about their own lives and opinions (see Appendix A for a complete list of prompts). All paragraphs were written and corrected using Google Sheets so that the students and rater could easily access their paragraphs at any time. Each student had their own document that they wrote and received feedback in. These also contained separate tabs where students could look to see the meaning of the various codes or to record the number of each error type they made in their first draft. An example of what these documents look like can be found in Appendix B. The first drafts of the first and last three prompts of each semester were grouped together to form the pre and posttest respectively. Only at the first draft was analyzed to determine whether student writing prior to feedback was affected by the treatment.

The control group was also given the first and last three prompts as a pre and posttest, but through two Google Forms surveys instead of Google Sheets. They were not given feedback of any kind until after they had completed the posttest and were not meant to review their pretest before taking their posttest, rendering the use of Google Sheets less practical in this instance.

Procedures

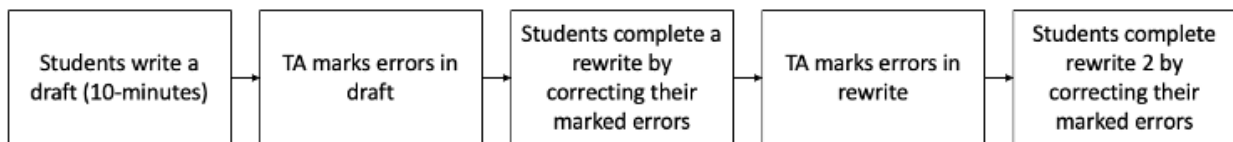
Towards the beginning of each semester, students in the treatment group were introduced to DWCF feedback and began to complete iterations of it using the 22 paragraph prompts. They typically completed their 10-minute paragraphs in class after a short, informal discussion on the topic. When writing the 10-minute paragraphs, students were encouraged to strive for the best grammar they could in response to these prompts. After completing these prompts, students were

given feedback in the form of coded corrections marking each mistake that the rater found. Students were then expected to rewrite their paragraphs and address as many of the errors as they could. The cycle was then repeated, with students given feedback using these same error markings and expected to complete a second rewrite (see Figure 1) As students were not expected to correct their final draft, these paragraphs were not marked or given feedback. By the end of the semester, students were expected to complete this cycle with all 22 prompts, excepting one semester where only 20 were required due to scheduling.

Like Hartshorn and Evans (2012) and Kurzer (2018), marked errors were divided into global, local, and mechanical errors. In this study, global errors were comprised of verb tense and form, sentence structure, word order, and transition word errors. Local errors included subject-verb agreement, determiner, singular/plural, count/noncount, word choice and form, and preposition errors. Finally, mechanical included spelling, punctuation, capitalization, missing words, unnecessary words, unclear meaning, and awkward phrasing. (A complete list and the codes for these errors can be found in Appendix B.) As noted in Kurzer (2018), these categories reflect whether an error is likely to distort meaning.

Figure 1

DWCF Process



For two semesters of the course, students received their feedback by the end of the day two classes after first writing their prompt (approximately one week excepting times that classes were canceled). These cycles overlapped, so a student might write their first draft for one prompt, second draft for another, and final draft for yet another all around the same time.

Students whose drafts or rewrites were not done in time also received feedback if they were turned in during the semester, although some that were turned in exceptionally late (such as during finals week) did not receive feedback.

Most errors, such as verb tense and punctuation errors, were marked with indirect, coded feedback for the first rewrite. For prepositions, students were given both the code and direct feedback on all drafts and rewrites. Sometimes additional errors received direct feedback based on the rater's understanding of whether the student would be able to correct the mistake on their own. For example, while many spelling errors were simply marked with the code, words that required students to use hyphens were often given direct feedback. This was done to ensure that students would understand that they needed to combine the two separate words they had written into a single word using a hyphen, instead of wondering what was wrong since the two words on their own were spelled correctly. Additionally, notes were occasionally left next to student paragraphs to help explain rules, what the best word choice would be, or to otherwise help, praise, or encourage the students. While this differs slightly from some DWCF studies, as noted in the delimitation section, this was part of the design of the writing course and done for student benefit.

Having only one rater for all student participants was done in part to help ensure intra-rater reliability. As this study was a pedagogical intervention using a university course, having multiple raters mark each paragraph and then resolve any conflicts before returning the paragraphs to the students in a timely manner was not feasible. Additionally, using error count to measure accuracy makes it more difficult to have agreement between raters when compared to error-free T-units or clauses. For these reasons, only a single rater was used. However intra-rater reliability of 89% agreement was obtained when 17 (approximately 9%) of the student essays

were re-coded 20 weeks later, suggesting high rater consistency. Agreeance was considered to be the same or comparable coding. For example, both “(D)” and “(^D)” were considered to agree as were codes that differed because there was more than one way to address the problem.

The control group originally received their pre and posttests in one of the university’s testing labs, but due to Covid-19 concerns and other circumstances, some students completed their posttest from a distance. Students were given instructions on how to respond to these prompts and asked to time themselves while completing their 10-minute paragraphs. An example of the instructions can be found in Appendix C.

Data Analysis

The collected data was analyzed for twelve different variables (see Table 2). First, the total number of errors in each of the error families were examined to assess DWCF’s effects on student accuracy. Several previous studies of DWCF have used error-free T-units or clauses compared to their errored counterparts to measure accuracy while the two most recent studies (Kurzer, 2018 and Eckstein et al., 2020) based their accuracy measures on error count (errors per 100 words and total normalized errors respectively). In light of these options, this study utilized normalized error scores for several reasons. First, they were used for the same reasons as Kurzer (2018) regarding the possibility of accuracy and fluency being measured together and increased specificity when looking at error type and count. While the original articles only addressed these drawbacks as being inherent to error-free T-units, they also apply to error-free clauses. Second, error-free T-units and classes also have the drawback of potentially masking errored writing if the same number of errors are distributed across T-units or clauses in one draft and then grouped into fewer T-units or clauses in another draft. Normalized errors are not sensitive to distribution within an essay. Finally, a .70 correlation has been found between measurements of error-free T-

units and errors per word (Polio & Shae, 2014). Thus, because error/word is a rather comparable measure, potentially mitigates any overestimations of student improvement, provides more detailed information about specific error types, and has been used in the most recent DWCF studies, it was chosen for this research.

Table 1

Variables

Accuracy		
		Global
		Local
		Mechanical
Lexical Complexity		
	LD	Lexical Density
	LS2	Lexical Sophistication (Lexical Frequency Profile)
	TTR	Type/Token Ratio
Syntactic Complexity		
	MLS	Mean length of sentence
	MLT	Mean length of T-unit
	CN/C	Complex nominal per clause
Fluency		
	W	Word count
	S	Sentence
	T	T-unit

Before analyzing the students' linguistic accuracy, the data had to be normalized. This study followed Biber et al (1998)'s normalization process (see also Ferris, 2001) by dividing each students' number of errors for their pre or posttest by the total number of words they wrote to ascertain their errors per word. This number was then multiplied by the average number of words from all students' pre and posttests in order to estimate the number of errors each student would have if they wrote something of average length. This normalized number was then used for all accuracy analyses. The normalization process was done multiple times so that each error family could be looked at separately.

For lexical and syntactic analysis, this study utilized lexical density, lexical sophistication, and type-token ratio for lexical analysis, which look at the ratio of content words, the number of high-frequency words, and how many unique words students use, allowing an examination of student word usage from a variety of angles. Syntactic analysis was measured using mean sentence length, mean length of T-unit, and complex nominals per clause, reflecting aspects of both phrasal and clausal sophistication (see also Biber and Gray, 2010). Number of words, sentences and T-units were chosen as fluency measures as they all measure different aspects of student output.

The lexical and syntactic measures match those of Eckstein et al. (2020), who argue that “These three [lexical] measures...illustrate writers’ breadth of open-class, low-frequency, and sophisticated word usage with the expectation that writers would show variety in these measures more readily than closed-class, high-frequency words” while the syntactic measures have been shown to reflect essay quality (2020, p. 89-90). The choice of this study’s variables is further supported by the College Board SAT Scoring Guide (2012) calling for “a varied, accurate, and apt vocabulary” (para. 2) and research which shows how L2 writers differ between L2 proficiency levels and when compared to their L1 counterparts in terms of complexity and fluency. This research indicates that L2 students may need help improving these aspects of their writing.

In a meta-study of 72 reports, Silva (1993) found that a majority of studies reported L2 writers as less fluent than L1 writers based on total word count. This research also reported also reported more but shorter T-units, fewer examples of noun modification, and “less lexical variety and sophistication” for L2 writers (p. 667). Abba (2015) reported that, when compared to their less proficient L2 counterparts, “Proficient L2 writers... demonstrated an increase in lexical and

syntactic diversity” (p. 17) including a wider range of words and an increased number of low-frequency words (see also Grant & Ginther, 2000; Crossley & McNamara, 2012; Crossley, Salsbury, & McNamara, 2012; Crossley, Salsbury, McNamara, & Jarvis, 2011). She also reported increased essay lengths among Generation 1.5 and L1 students when compared to L2 writers and differences in syntactic complexity, use of high-frequency words, lexical diversity, and lexical sophistication between L1 and L2 writers (see also Crossley & McNamara, 2009; Di Gennaro, 2009; Kormos, 2011). The differences between L2 writers and their peers suggest areas of L2 student writing that should be examined carefully to see under what circumstances students show improvement in these areas.

In sum, because these accuracy, complexity, and fluency measures reflect established criteria for measuring changes in L2 student writing and examine specific areas of concern for these writers, they were included as this study’s variables to evaluate changes among the control and treatment groups. These syntactic and lexical complexity measures were respectively analyzed using Ai and Lu’s Web-based L2 Syntactical Complexity Analyzer and Web-based Lexical Complexity Analyzer (Ai & Lu, 2010; Ai & Lu, 2013; Lu, 2010, 2011, 2012; Lu & Ai, 2015).

As many variables were being analyzed all at once, a Bonferroni adjustment was used to prevent false positives (Type I errors). After dividing the typical p-value threshold for significance of .05 by the number of tests, the results of this study were determined to be significant only if they had a p-value of .004 or lower.

After the data was normalized and the complexity analyses completed, the treatment and control group were compared to see if there was a statistically significant difference in each group between their pre and posttests. This was done using two-way repeated measures (RM)

ANOVAs, which examine both the effects of time and the effects of treatment on the students' writing.

CHAPTER 4: RESULTS

Research Question 1: Accuracy

Research question 1 looked at whether student accuracy changed after 14-weeks of DWCF treatment, specifically how the error families changed compared to each other. After conducting RM ANOVAs for all error families, there were no significant results for global, $F(1, 31) < 0.001, p = 0.986$, local, $F(1, 31) = 0.004, p = 0.950$, or mechanical, $F(1, 31) = 0.131, p = 0.720$, errors despite both groups seeing a decrease in errors over time. While the treatment group had higher numbers of errors on both the pre and posttests across all error families, they decreased a comparable amount relative to the control group (refer to Table 3 for more detail).

The initial prediction that errors would decrease after treatment, but potentially be insignificant, was supported by the data. However, the prediction that some error families would have differing results is only true on an insignificant level.

Table 3

		<i>Accuracy Results</i>						
		N	M		SD		p	η^2p
			Pre	Post	Pre	Post		
Global	Control	9	5.760	4.450	4.850	2.330	0.986	0.000
	Treatment	23	8.720	7.450	6.080	4.930		
Local	Control	9	15.000	10.500	9.030	4.880	0.950	0.000
	Treatment	23	26.000	21.900	22.200	12.700		
Mechanical	Control	9	20.000	14.800	10.600	8.560	0.720	0.004
	Treatment	23	29.000	26.500	23.200	17.100		

Figure 2

Global Errors

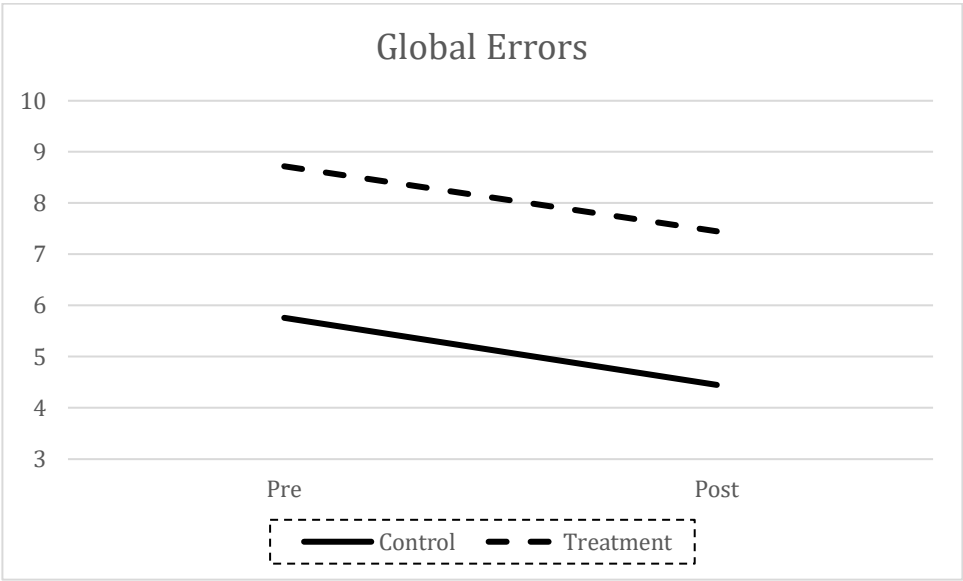


Figure 3

Local Errors

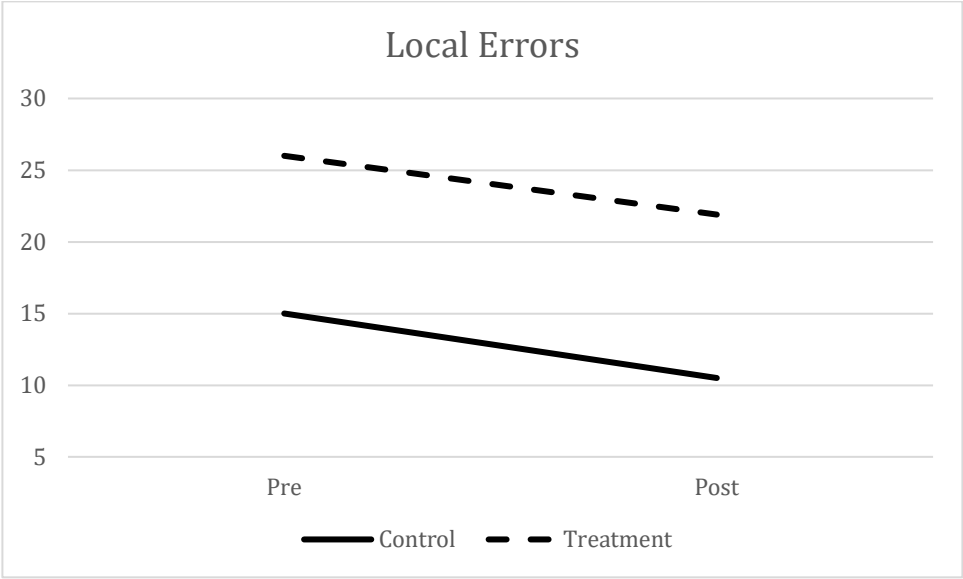
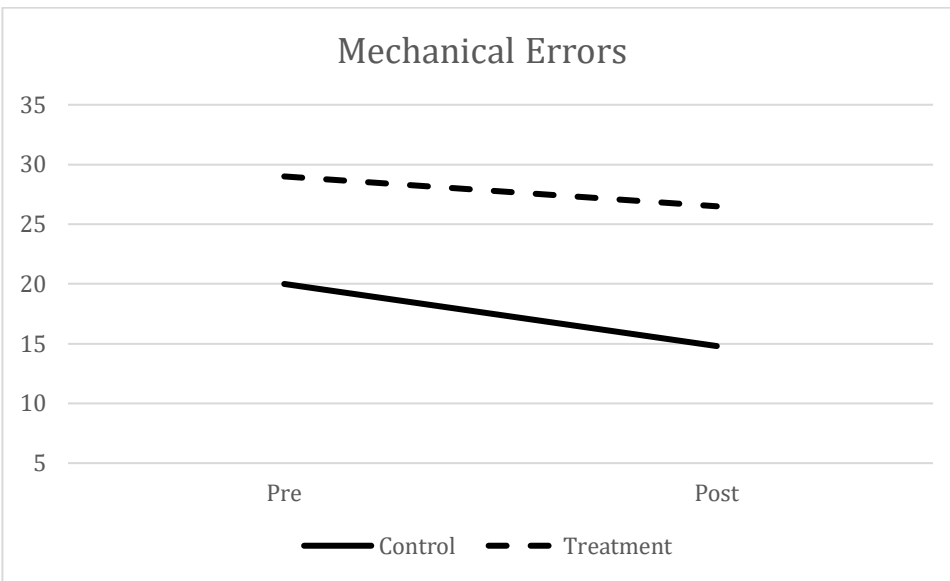


Figure 4

Errors Mechanical



Research Question 2: Complexity

Research question 2 inquired how syntactic and lexical complexity measures would change after a 14-week DWCF treatment. RM ANOVAs were again used for each measure individually. With regard to the syntactic complexity measures, all measures—mean length per sentence, mean length per T-unit, and complex nominals per clause—decreased over time for both groups (see Table 4). None had significant results, MLS, $F(1, 31)=0.038, p=0.848$, MLT, $F(1, 31)=0.136, p=0.715$, and CN/C, $F(1, 31)=1.000, p=0.325$, meaning that there were no statistically significant effects between the two groups.

Table 4

Syntactic Complexity Results

		N	M		SD		p	η^2p
			Pre	Post	Pre	Post		
MLS	Control	9	20.000	17.400	3.550	2.880	0.848	0.001
	Treatment	23	19.000	16.700	4.110	2.930		
MLT	Control	9	16.900	15.200	2.330	1.550	0.715	0.005
	Treatment	23	15.900	14.700	3.120	2.790		
CN/C	Control	9	1.110	0.773	0.217	0.166	0.325	0.032
	Treatment	23	0.222	0.201	0.222	0.201		

Syntactic complexity: Mean length of sentence (MLS), mean length of T-unit (MLT), and complex nominals per clause (CN/C).

Figure 5

Mean Length of Sentence

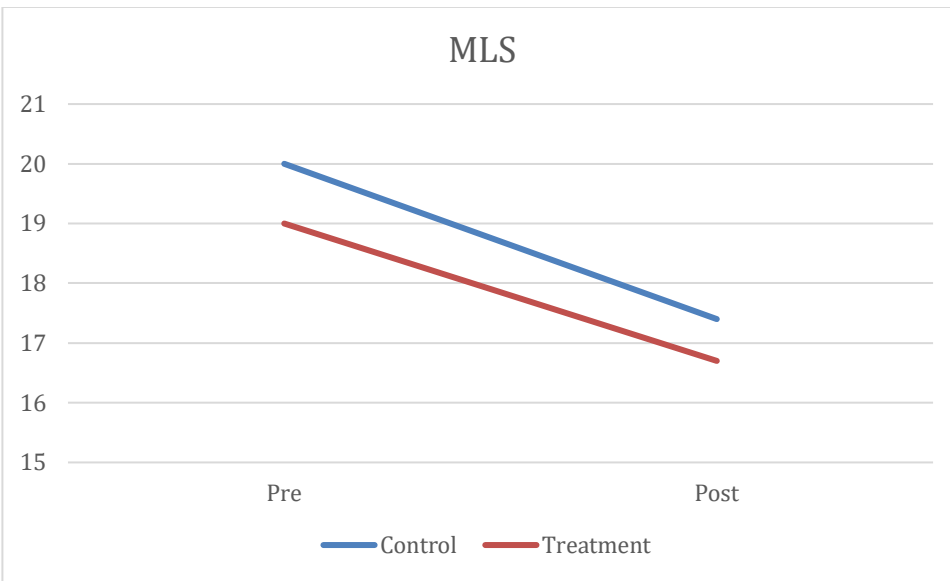


Figure 6

Mean Length of T-unit

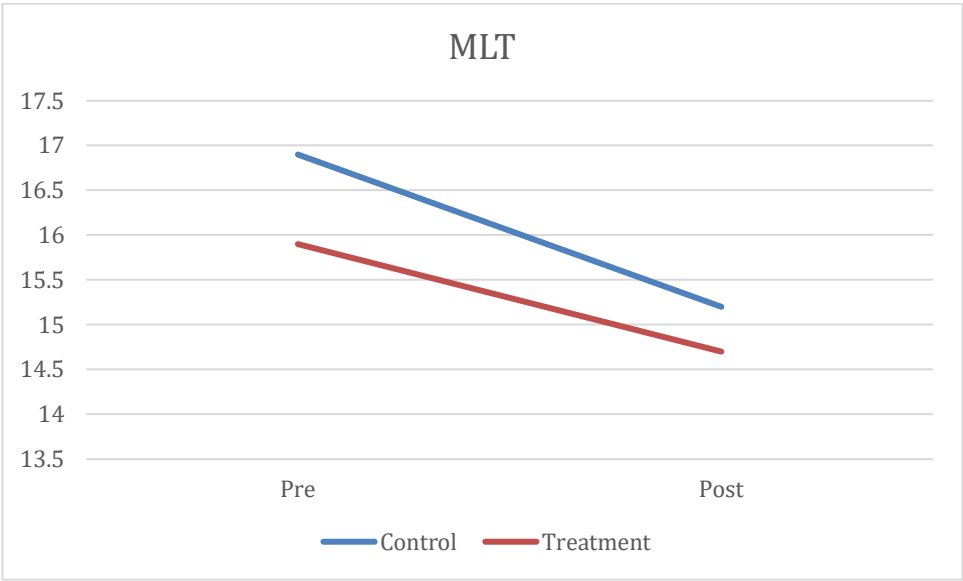
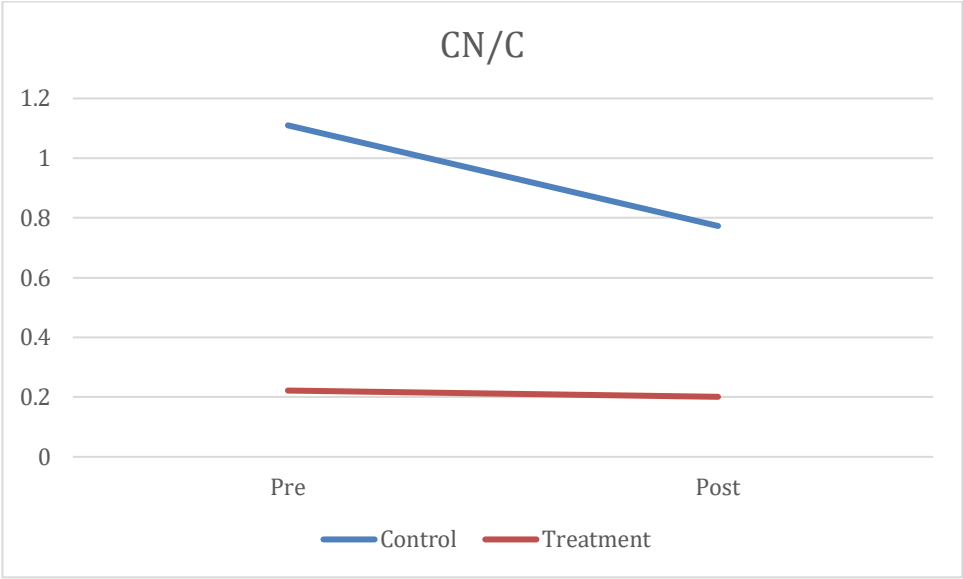


Figure 7

Complex Nominals per Clause



For the lexical complexity measures, there were no significant results for lexical density ($F(1, 31)=4.970, p=0.777$), lexical sophistication ($F(1, 31)=<0.001, p=0.978$), or type-token ratio ($F(1, 31)=0.446, p=0.509$). Although both groups showed a decrease in lexical density and lexical sophistication between their pre and posttests, the treatment group showed a decrease in type-token ratio while the control group's type-token ratio increased (see Table 5).

Table 5

Lexical Complexity Results

		N	M		SD		p	η^2p
			Pre	Post	Pre	Post		
LD	Control	9	0.512	0.498	0.038	0.022	0.777	0.003
	Treatment	23	0.516	0.497	0.028	0.033		
LS2	Control	9	0.228	0.203	0.037	0.022	0.978	0.000
	Treatment	23	0.241	0.217	0.044	0.042		
TTR	Control	9	0.448	0.464	0.074	0.104	0.509	0.015
	Treatment	23	0.442	0.441	0.066	0.058		

Lexical complexity: Lexical density (LD), lexical sophistication (LS2), and type-token ratio. (TTR)

Figure 2

Lexical Density

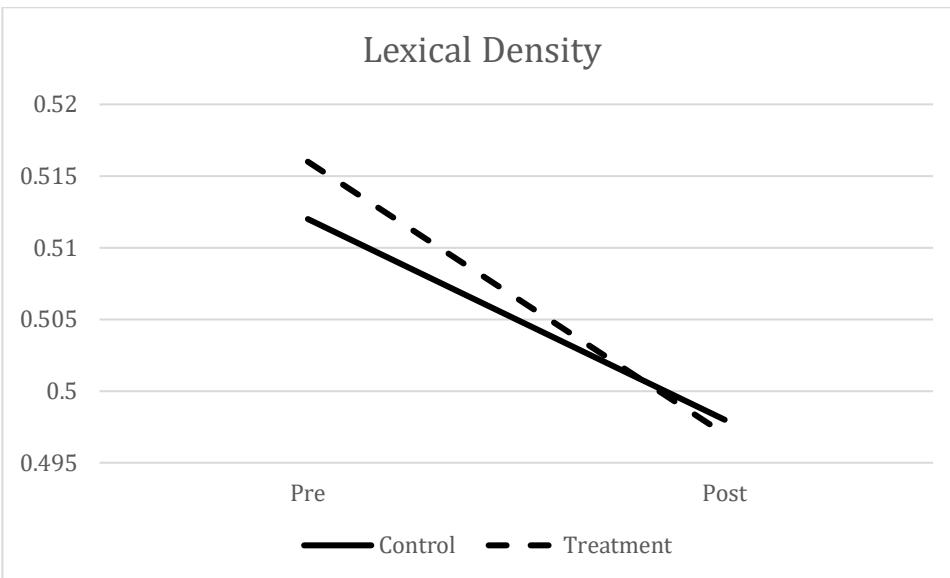


Figure 3

Lexical Sophistication

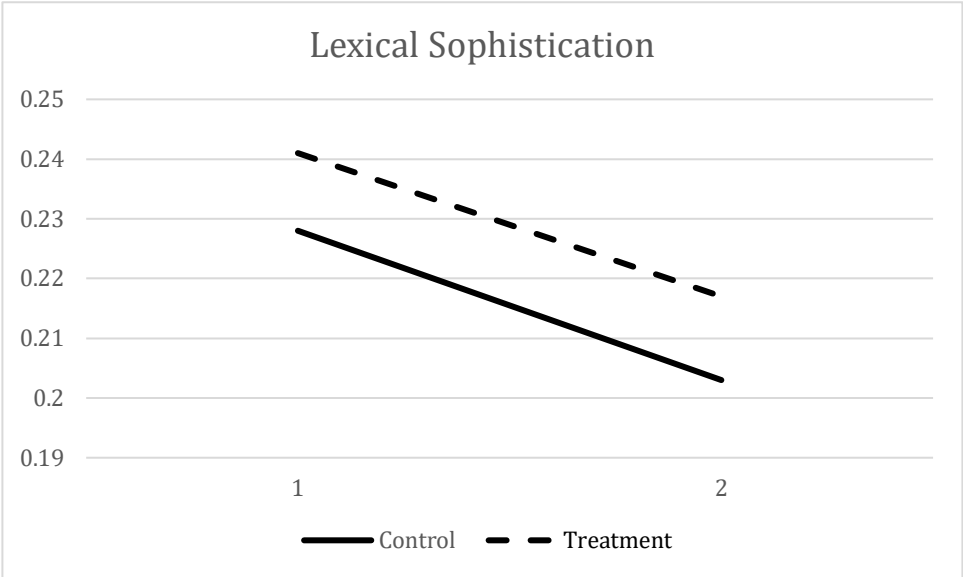
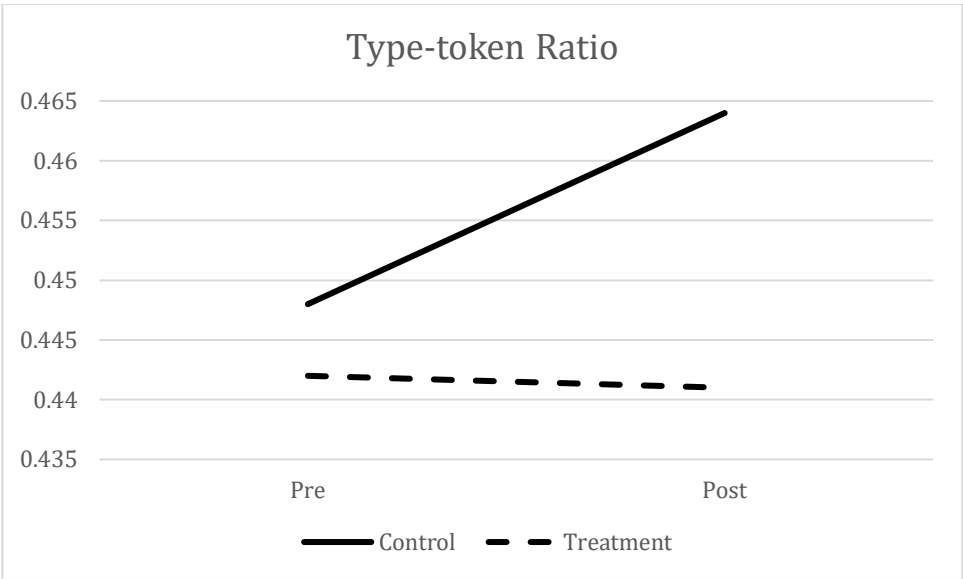


Figure 4

Type-token Ratio



The initial hypothesis that neither syntactic nor lexical measures would be significantly affected by the treatment was supported by this data.

Research Question 3: Fluency

The final research question looked at how writing fluency would be affected by DWCF. In regard to writing fluency, again there were no significant results for word count ($F(1, 31) < 0.000, p = 0.985$), sentence count ($F(1, 31) = 0.004, p = 0.950$), or T-units ($F(1, 31) = 0.039, p = 0.845$) after separate RM ANOVAs were performed. For both groups of students, their sentence numbers increased between the pre and posttest, but with no noticeable difference between the groups. In contrast, student word count decreased in both groups over time. Finally, while there were no significant differences between the treatment and control groups, the treatment group slightly increased their T-unit count and the control group decreased slightly (see Table 6).

The hypothesis for this research question posited that writing fluency would either be unaffected or decrease after treatment due to student focus on grammar. The insignificant results support this hypothesis, but the initial hypothesis does not fully account for the changes (though insignificant) that did occur as they occurred in both groups.

Table 6

Fluency Results

		N	M		SD		p	η^2p
			Pre	Post	Pre	Post		
Word	Control	9	515	470	159	195	0.985	0.000
	Treatment	23	514	470	148	135		
Sentence	Control	9	25.900	27.300	7.440	10.800	0.950	0.000
	Treatment	23	27.600	28.800	8.270	9.250		
T-unit	Control	9	30.600	31.200	8.800	13.000	0.845	0.001
	Treatment	23	33.300	33.000	10.900	11.000		

Fluency: Total words, total sentences, and total T-units.

Figure 5

Word Count

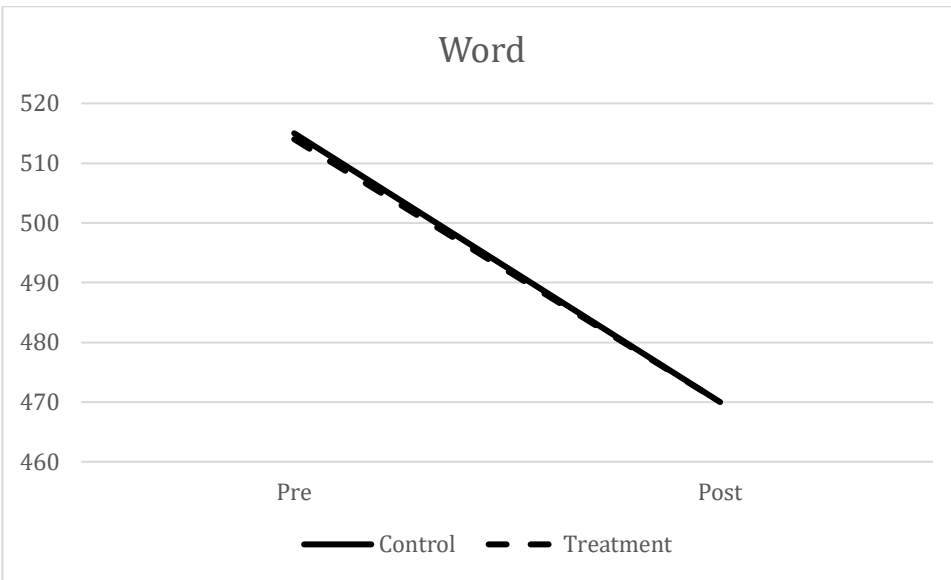


Figure 6

Sentence Count

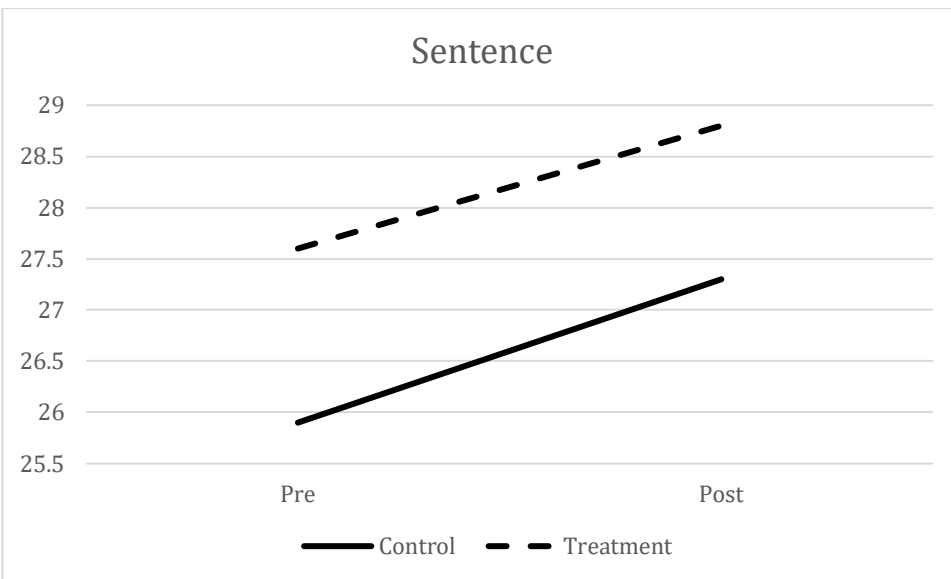
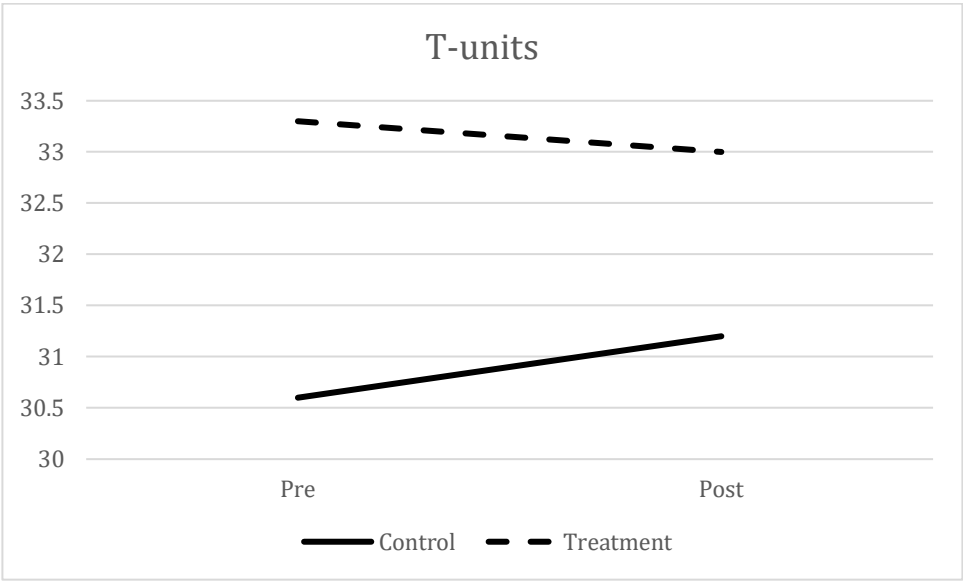


Figure 7

T-unit Count



CHAPTER 5: DISCUSSION AND CONCLUSION

Discussion

Research Question 1

The first research question for this study asked what effect a 14-week DWCF intervention has on L2 graduate students' linguistic accuracy in terms of global, local, and mechanical error families. The expected result was that student errors would reduce, significantly or not, and that the effects on each error family would noticeably differ. The findings only partially support this hypothesis as all error types were reduced for both treatment groups and there was no significant difference between error families, which all decreased with no significant differences between them.

These results contrast with the findings of DWCF studies on IEP and undergraduate students (Hartshorn et al., 2010; Evans et al., 2010; Lee, 2009; Evans et al., 2011; Hartshorn & Evans, 2015), but shared similar results to other studies with graduate level students (Eckstein et al, 2020) suggesting the possibility of a ceiling effect on accuracy gains when using DWCF with high-proficiency students.

Additionally, the decrease among all error families, while insignificant in this study, matches most closely with the results of Kurzer (2018) as well as the timely feedback group in the study by Eckstein et. al. (2020) rather than with Hartshorn & Evans (2012) findings that only some error groups had significant changes. All four studies used comparable error categories of global, local, and mechanical errors. Together, most evidence suggests that all error families will be affected by a DWCF treatment in the same way (i.e., all showing significant or insignificant effects), although there is still some possibility of only some categories having significant effects. With this in mind, it is likely that IEP and undergraduate students who struggle with any

type of accuracy errors will benefit from DWCF, while graduate students who struggle are likely to see improvement, but not to the same extent.

The results of this study regarding accuracy may be due to the high proficiency level of these graduate students. In other words, there is a high likelihood that DWCF's usefulness is limited by a ceiling effect as hypothesized by Kurzer (2018) and supported by Eckstein et al. (2020). It could be that with their high proficiency level, graduate students do not make enough errors to show significant improvement regardless of what instructional strategy is used. It is also possible that the decrease in both groups was due to these highly motivated L2 students making efforts to improve on their own; improving due to other courses in their studies; or learning from working on theses, dissertations, and other articles. The continual demands of graduate school and life in an English-speaking country may also have forced both groups to improve over time. It is possible that it was simply a natural result of being immersed in schooling and research in their L2; perhaps many students received other types of feedback on their writing that had the same benefit as DWCF. While the reasons are not certain at this time, thus far this is the only research that has used a control group to look at graduate student accuracy and results do not indicate that DWCF has any significant effects on these students.

Research Question 2

Research question 2 focused on DWCF's effects on syntactic and lexical complexity, predicting that there would be no significant effects on any of the measures. While all results were insignificant, not all changes in lexical complexity showed the same effect. While lexical density and sophistication decreased over time for both groups, type-token ratio increased only in the control group. On the other hand, all syntactic complexity measures in both groups decreased

but were likewise not statistically significant. This lack of significance in all complexity measures suggests that the effects may not have been caused by the treatment.

The lexical and syntactic complexity results closely mirror the findings of the research done by Eckstein et al. (2020) with the exception that Eckstein et al.'s timely feedback group showed an increase in TTR while only their postponed group reflected a decrease like found in this study's control group. However, as there were no significant results across any of these categories, this study seems to correlate best with the findings of Hartshorn & Evans (2015) of no significant effects on complexity after DWCF treatment. As such, students whose main writing concern is a lack lexical complexity may not see the desired changes in their writing through DWCF treatment.

The insignificant changes in lexical complexity measures have several possible explanations. A possible explanation for the changes in TTR is that spelling errors may have affected lexical diversity, but the lack of significant changes in mechanical errors makes this less likely although not impossible as this was found to be the cause in research done by Eckstein et al. (2020). Further analysis of the data would be required to confirm whether this is the case for this study's results. Regardless, the lack of significant effects does not provide evidence that graduate students are lexically affected by DWCF.

The syntactic complexity results showed all measures decreasing over time for both groups. While much of the research on syntactic complexity has looked at how increased clauses indicate increased clausal complexity, Biber and Gray (2010) have argued that phrasal complexity is reflected by compression in academic writing. Thus, the reduction in syntactic complexity measures may actually reflect some increases in phrasal complexity. However, to be certain whether this study's results reflect an increase, decrease, or mixed effects on complexity

would require corpus tagging. Corpus research has not been used up to this point in DWCF, but the results of this study and other studies looking at writing complexity suggest a need for it. At this time, however, the insignificant results of this study do not provide evidence that DWCF affects graduate student syntactic complexity.

Finally, the results of both lexical and syntactic complexity provide limited support for the trade-off hypothesis (Skehan, 1998) with (insignificant) accuracy gains and overall complexity losses among both groups.

Research Question 3

The final research question addressed writing fluency with the prediction that there would either be no effect or a reduction in fluency as students focused on grammatical accuracy. While the latter could seem to be the case when looking only at the treatment group's decreased word count, the control group experienced decreased word counts as well and neither groups' change was significant. This suggests that the former prediction, no noticeable effect, was most accurate. Sentence count also showed similar trajectories in both groups, although for both groups the sentence count increased. These changes were also insignificant, again suggesting no effect from DWCF on student fluency. Finally, although insignificant, the two groups differed when looking at total T-unit counts with an increase of T-units in the control group and a decrease among the treatment group.

These fluency results contradict the findings of Evans et al. (2011) and Hartshorn et al. (2010) who found that DWCF created (insignificant) decreases across all fluency measures. Instead, it matches up more closely with Eckstein et al. (2020) who found a mix of decreased and increased fluency measures. However, unlike Eckstein et al., none of the findings from this research were significant. The lack of evidence provided by this research seems to indicate that

the likelihood of DWCF having significant effects on fluency is low and likely to occur only in certain measures. In sum, DWCF is unlikely to provide students with a fluency benefits but is also unlikely to cause such significant detriments to fluency to be avoided only for that reason.

If DWCF was the cause of these insignificant changes, there are several possibilities for why they were affected in these ways. The shared reduction of word counts again provides weak support for the trade-off hypothesis (Skehan, 1998) as both groups improved in accuracy while reducing their overall fluency. An optimistic view of the reduction of T-units would be that the control group may have learned not to use extraneous words or run on sentences, resulting in more concise academic writing, an aspect of phrasal complexity as argued by Biber and Gray (2010). This could also explain why treatment group students saw a reduction in overall word count but would not explain why the control group saw a similar reduction unless both groups were developing more phrasal complexity regardless of DWCF treatment. For now, the results simply provide no evidence of DWCF effects on graduate student fluency.

Discussion Summary

While there is still much to be learned about DWCF, this study supports several previous studies that suggest it has little effect on areas of writing outside of written accuracy (Hartshorn et al., 2010; Hartshorn & Evans, 2015; Eckstein et al., 2020). In addition, this research suggests that there is no significant impact on student accuracy gains for L2 graduate students. If this lack of impact is due to the students' high proficiency level, it is likely that DWCF will also not benefit the accuracy of those learners who are above this accuracy level either.

The lack of statistical significance among all measures remains true even when compared to the pre-Bonferroni adjustment significance threshold of .05. While Bonferroni adjustments can lead to Type II errors (false negatives), these results indicate that it is not the case for this

study. In sum, the findings of this study reflect very little correlation between accuracy, complexity, and fluency measures and DWCF even with a less conservative statistical approach, which strengthens the evidence that these measures are not significantly affected by DWCF at the graduate student level.

Delimitation and Suggestions for Future Research

Some in the composition field would argue that by not giving an overall score or a rhetoric score that this study did not look at writing, but simply practice with language. The writing course used for this study was designed to fit easily into any graduate semester without detracting from the students' ability to focus on their chosen field and research. Almost all assignments were intended to become part of a paper, thesis, or dissertation that would be used outside of the class as well, with other assignments being largely for practice (and research purposes), including the DWCF paragraphs. As these were meant as grammar practice and instruction within the context of the course and the course itself was meant to allow students to learn and grow without the stress that a non-major specific class could bring down their GPA and affect future scholarship, students were not graded on their DWCF paragraphs beyond completion. We acknowledge that, along with the findings of other DWCF studies that do not look at overall scores, these findings may not be transferable to process writing or other types of writing that differ from those used in the treatment group (see also Kurzer, 2018). However, there is a high probability of transference based on skill acquisition theory (see DeKeyser, 2007) and the findings of Akiyama and Fleshler (2013), whose DWCF students showed accuracy improvement on spontaneous essay writing. Future DWCF researchers are encouraged to look further at the connection between DWCF and outside writing.

Some deviations to the process of DWCF itself may have affected the outcomes of this study. Having some feedback be coded—typically either because students had not yet self-corrected it properly and had only one draft left or because it was unlikely they would be able to—again reflects the course itself and the desire to benefit these students to the greatest degree. Students who were not able to self-correct immediately would still be able to see the correct answers and learn from their errors after trying to succeed on their own. Additionally, students would not waste time (as, once again, it was important that the course benefit and not burden the students) looking up something they were unlikely to correct without help; the help was simply provided in advance. While a future comparison of this method and the original DWCF procedures could help determine the likelihood of this affecting this study’s results, double-rated essays may be preferable to allow for more robust findings. Additionally, future studies could examine various ratios of uncoded feedback cycles to coded feedback cycles to see what, if any, effects they have on DWCF results.

Another aspect of note is the limited number of cycles used in this study. This also reflects a desire to avoid overburdening the students and, additionally, was hypothesized to be less important for higher-level students, who were expected to make fewer errors than students of lower proficiency levels. Additionally, we felt that many educators will similarly adapt DWCF in some way for their classroom and that this research is likely to show the results of DWCF being adapted to individuals and their classes while still reflecting many of the key tenets of the method itself. However, with this in mind, it is essential that future studies using graduate students explore whether or not this change is the reason for the insignificant results. Should future studies find that no amount of drafting produces significant results at the graduate level, then the current theory that there may be a ceiling effect becomes more likely. Should other

research yield significant results when graduate students are given more drafts, then there is evidence that too many changes to the original DWCF model, even well-intentioned changes, may impact this strategy's effectiveness. As such, we encourage future researchers to continue to look into research both with graduate level students and with variations in draft number to fully flesh out an understanding of what impacts these may have had.

Third, many previous studies on DWCF have used error-free T-units or error-free clauses to measure accuracy. Future research analyzing the same data (from this study or otherwise) using both error count and error-free T-units could provide insight into their comparability and the accuracy of Polio and Shae (2014)'s correlation estimate between the two when research data is used for DWCF.

Finally, it should be noted that no previous DWCF studies reported in-class discussions of the topic before students wrote their paragraphs. This allowed treatment group students more time to think about the topic and potentially provided them with some of the words and phrases they used. With no significant difference between groups in this study, it may be that the effects of such a discussion are too small to make a noticeable difference, but this can only be certain if future research compares groups who do and do not discuss the topic before writing.

In addition to changes to DWCF, another possible limitation is that, as opposed to having too high of an English-speaking level for DWCF to affect them, these students possess different motivations than students at the IEP and undergraduate levels. Perhaps knowing that they have already entered their program of choice or the knowledge that someone can edit their research before publishing has led to a decrease in student motivation regarding their English grammar. Previously, most DWCF research was done in high-motivation ESL contexts—IEP and undergraduate students—and future research may benefit from exploring other contexts where

motivation may differ, including both continued research with graduate students and research into EFL students, whom Ferris (1999) notes may have significantly different motivations from ESL students. It is quite possible that student goals and motivations alter the effectiveness of this method, leaving a wide range of research potential regarding student motivation and DWCF.

One final limitation of this study is the limited number of students in the control group and the disciplines that students from both groups belonged to. By itself, a small control group provides weaker evidence than a more robust group. Additionally, writing expectations differ from department to department, with fields such as computer science and engineering requiring much different expectations than other departments (Hartshorn & Evans, 2019). It is possible that this creates a range of proficiency levels among students, with all having a baseline proficiency that allowed them to pass an English proficiency test and the GRE or GMAT (if required), but not all passing with as high of scores. This may also affect the types of feedback students received outside of the linguistics class and their personal focus and motivations regarding writing and accuracy improvement.

Pedagogical Implications

Although this research did not find any evidence of differences between L2 graduate level students who received DWCF and their peers in the control group, it also has also not definitively shown that DWCF is only beneficial to accuracy below the graduate level. Based on this and past research, it can be hypothesized that DWCF benefits student accuracy for intermediate level students and some advanced students but may cease to provide significant accuracy improvements once students become highly advanced. This hypothesis would require further studies, including studies that compare graduate and undergraduate students, in order to fully explore this possibility. In regard to complexity and fluency, the results of this study

indicate that DWCF is unlikely to significantly increase or decrease complexity and fluency measures, supporting some of the current research (Hartshorn & Evans, 2015) and in contrast to others (Eckstein et al., 2020).

While educators of L2 undergraduate and pre-matriculated students have multiple studies to support DWCF's accuracy impact on their students and can be relatively confident of its accuracy effects on their students, graduate-level instructors have only limited evidence at this point in the research. The question of whether their students' accuracy improves regardless of DWCF is still uncertain. Both this and the still uncertain effects on complexity and fluency should be factored into graduate-level educators' decision whether to use DWCF while further research is being conducted.

Conclusion

While indicated to be effective at lower levels, the effects of DWCF on L2 graduate students has, thus far, not shown evidence of having the same degree of effect on graduate students' written accuracy. As all other results are largely insignificant, they suggest that DWCF may have some influence on L2 graduate students' writing complexity and fluency, but there is not yet sufficient evidence to be certain of these effects. We hope that with this knowledge in mind, graduate-level educators will be able to make better-informed choices for their students while anticipating further research into this area.

References

- Ai, H. and Lu, X. (2010). *A web-based system for automatic measurement of lexical complexity*. Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10). Amherst, MA. June 8-12.
- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. *Automatic treatment and analysis of learner corpus data*, 249-264.
- Akiyama, Y., & Fleshler, M. (2013). Effects of dynamic written corrective feedback on JFL students' homework. *New perspectives on Japanese language learning, linguistics, and culture*, 49-68.
- Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of experimental psychology: learning, memory, and cognition*, 23(4), 932.
- Bates, L. (1993). *Writing Clearly: Responding to ESL Compositions*. Heinle and Heinle Publishers, Wadsworth Book Distribution Center, 7625 Empire Drive, Florence, KY 41042.
- Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2-20
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Bitchener, J., & Knoch, U. (2015). Written corrective feedback studies: Approximate replication

- of Bitchener & Knoch (2010a) and Van Beuningen, De Jong & Kuiken (2012). *Language Teaching*, 48(3), 405-414.
- Burt, M. K., & Kiparsky, C. (1972). The Gooficon: A Repair Manual for English.
- BYU Graduate Studies. (2021a). *English Proficiency*.
<https://gradstudies.byu.edu/page/english-proficiency>
- BYU Graduate Studies. (2021b). *Admissions Guide*.
<https://gradstudies.byu.edu/page/admissions-guide>
- Chomsky, N. (2002). *Syntactic structures*. Walter de Gruyter.
- Connors, R. J., & Lunsford, A. A. (1988). Frequency of formal errors in current college writing, or Ma and Pa Kettle do research. *College composition and communication*, 39(4), 395-409
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18(2), 119-135.
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115-135.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, 29(2), 243-263
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561-580.
- DeKeyser, R., VanPatten, B., & Williams, J. (2007). Skill acquisition theory. *Theories in second language acquisition: An introduction*, 97113.

- Di Gennaro, K. (2009). Investigating differences in the writing performance of international and Generation 1.5 students. *Language Testing*, 26(4), 533-559.
- Doughty, C., & Varela, E. (1998). Communicative focus on form. *Focus on form in classroom second language acquisition*, 1, 114-138.
- Doughty, C., & Williams, J. (1998). *Focus on Form in Classroom Second Language Acquisition. The Cambridge Applied Linguistics Series*. Cambridge University Press, 40 West 20th Street, New York, NY 10011-4211 (paperback: ISBN-0-521-62551-3; hardcover: ISBN-0-521-62390-1).
- Eckstein, G. (2018a). Goals for a writing center tutorial: Differences among native, non-native, and generation 1.5 writers. *WLN: A Journal of Writing Center Scholarship*, 42(7-8), 17-24
- Eckstein, G. (2018b). Keeping Good Company: How Chomsky's Universal Grammar Supports Truscott's Advice to Abandon Grammar Correction in Writing Instruction. *Language Teaching*, 38(02), 81-89.
- Eckstein, G., Sims, M., & Rohm, L. (2020). Dynamic Written Corrective Feedback among Graduate Students: The Effects of Feedback Timing. *TESL Canada Journal*, 37(2), 78-102.
- Ellis, R. (1998). Teaching and research: Options in grammar teaching. *TESOL Quarterly*, 32(1), 39-60.
- Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research*, 14(4), 445-463.
- Evans, N. W., James Hartshorn, K., & Strong-Krause, D. (2011). The efficacy of dynamic

- written corrective feedback for university-matriculated ESL learners. *System*, 39(2), 229-239.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of second language writing*, 8(1), 1-11.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be?. *Journal of second language writing*, 10(3), 161-184.
- 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(1), 49-62.
- Ferris, D. (2009). Theory, Research, and Practice in Written Corrective Feedback: Bridging the Gap, or Crossing the Chasm?. *New Zealand Studies in Applied Linguistics*, 15(1), 1.
- Gillespie, P., & Lerner, N. (2000). *The Allyn and Bacon Guide to Peer Tutoring*. Allyn and Bacon, 160 Gould St., Needham Heights, MA 02494.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of second language writing*, 9(2), 123-145.
- Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, Diane., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *Tesol Quarterly*, 44(1), 84-109.
- Hartshorn, K. J., & Evans, N. W. (2012). The differential effects of comprehensive corrective feedback on L2 writing accuracy. *Journal of Linguistics and Language Teaching*, 3(2), 217-248.
- Hartshorn, K. J., & Evans, N. W. (2019). Expectations and Challenges of Non-native University Writers at the Outset of Discipline-specific Study. *TESL Reporter*, 52(1).

- Hendrickson, J. M. (1980). The treatment of error in written work. *The Modern Language Journal*, 64(2), 216-221
- James, C. (1998). Errors in language learning and use: Exploring error analysis. London: Longman.
- 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. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2), 148-161.
- Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL Quarterly*, 52(1), 5-33.
- Lane, J., & Lange, E. (1999). *Writing clearly: An editing guide*. Boston: Heinle & Heinle.
- Lane, J., & Lange, E. (2012). *Writing clearly: Grammar for editing* (3rd ed.). Boston, MA: Heinle Cengage Learning.
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, 30(4), 579-589
- Lee, S. (2009). The efficacy of dynamic written corrective feedback on intermediate-high learners' writing accuracy. Unpublished master's thesis, Brigham Young University, Provo, Utah, USA.
- Lightbown, P. (1998) The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom SLA*. Cambridge.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of

- college-level ESL writers's language development. *TESOL Quarterly*, 45(1), 36-62.
- Lu, X. (2012). The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives. *The Modern Language Journal*, 96(2), 190-208.
- Lu, X. & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in Communicative classrooms. *Studies in Second Language Acquisition*, 19, 37–66.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. *Cognition and second language instruction*, 288
- Robinson, P. (2003). The cognitive hypothesis of adult, task-based language learning. *Second Language Studies*, 21, 45-107
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL-International Review of Applied Linguistics in Language Teaching*, 43(1), 1-32
- Robinson, P. (2007a). Criteria for classifying and sequencing pedagogic tasks. *Investigating tasks in formal language learning*, 7, 26
- Robinson, P., & Gilabert, R. (2007b). Task complexity, the cognition hypothesis and second language learning and performance. *IRAL: International Review of Applied Linguistics in Language Teaching*, 45(3).
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press
- Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors. *Studies in*

Second Language Acquisition, 11, 385–395.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327-369.

Truscott, J. (1999). The case for "the case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, 8(2), 111-122.

Truscott, J., & Hsu, A. Y. P. (2008). Error correction, revision, and learning. *Journal of second language writing*, 17(4), 292-305.

Appendix A

List of Prompts

1	Describe the most difficult course you have ever taken.
2	Name a controversial topic in your field and delineate your position.
3	Why did you choose your area of study?
4	Why did you choose to attend BYU?
5	Why is writing an important skill in the modern world?
6	Which has a greater impact on performance: hard work or talent?
7	Is it better to be too busy or too free?
8	Explain what makes art appealing (or not appealing) to you?
9	Is music truly a universal language? Explain.
10	Should a college education be free? Why or why not?
11	As a whole, does the internet help or harm society? Are there any exceptions?
12	In your opinion, what is the best study environment?
13	What is the most difficult aspect of being religious in the modern world?
14	Is there ever a time where plagiarism is morally acceptable? Please explain.
15	Are smaller families better than larger ones? Why or why not?
16	What is the most valuable job for society? Has this ever changed?
17	Under what circumstances is killing justified, if at all?
18	Name three reasons why the government should or should not ban certain firearms.
19	Support your position on veganism and other alternative diets.
20	Is there a superior pet? What is it and why?
21	Is bullying an issue that should be addressed by schools or left to parents?
22	According to a Czechoslovakian proverb, "Better a lie that soothes than a truth that hurts." Do you agree?

Appendix B

Treatment Group Materials

Figure A1

Student Paragraph Examples

Paragraph Revisions	
1	
2	Directions: During class you will write a paragraph for 10 minutes in the "draft" cell. Your TA will mark the errors; correction symbols will be placed after the word, phrase, or sentence that is errored. Then, copy and paste your paragraph into the "rewrite" cell and make the needed corrections. Usually a second rewrite will not be required.
3	Other prompt:
4	Prompt 1: Describe the most difficult course you have ever taken.
5	Draft One of the most difficult classes that I remember I(X) ever(X) took(VT/VF) was "[Course Name]".(P) I took it as an elective during my masters(P). My seniors informed be(SPG) before that the class is hard to pass. I didn't have any other option other than taking the course and also(P) I love physics(,) which made me choose the class. The reason why the class is hard because of the way you get your grade.(SS) There is only one final three hour examination at the end of semester which determines your grade. It got(WC) six questions with multiple sub questions(SPG sub-questions). All sub questions(SPG) are connected in the sense,(P/^) if you can't answer the first sub question(P/SPG) you cannot answer any of the other remaining sub questions(SPG). Most of the questions are complicated problems which took me about an hour to solve. I was required to finish six questions within three hours. I was slow and couldn't finish more than three in my first attempt. I failed the first time I took the examination. Almost more than half ^ failed the course and retook it. The next semester, I studied and practiced really hard. I passed the course but not with a(D) grade I expected. One of my other international could not pass the test (D) second time and he had to change his masters.(P) At(PP in) the end, I can say taking the course was a great experience for me. I learned a lot.
6	Rewrite One of the most difficult classes that I remember taking was [Course Name]. I took it as an elective during my master's.(SS) My seniors informed me before that the class is hard to pass. I didn't have any other option other than taking the course and also, I love physics(P) which made me choose the class. The way we get the final grade is the reason why the class is hard. There is only one final three hour examination at the end of semester which determines your grade. It has six questions with multiple sub-questions. All sub-questions are connected in the sense that if you can't answer the first sub-question, you cannot answer any of the other remaining sub-questions. Most of the questions are complicated problems which took me about an hour to solve. I was required to finish six questions within three hours. I was slow and couldn't finish more than three in my first attempt. I failed the first time I took the examination. Almost more than half ^ had failed(VF) the course and retook it. The next semester, I studied and practiced really hard. I passed the course but not with the grade I expected. One of my other international friends could not pass the test the second time and he had to change his masters(P). In the end, I can say ("that [grammatically correct both ways, but more formal with "that"]) taking the course was a great experience for me. I learned a lot.
7	Rewrite 2 One of the most difficult classes that I remember taking was [Course Name]. I took it as an elective during my master's. Our seniors informed us before that the class is hard to pass. I didn't have any other option other than taking the course and also, I love physics, which made me choose the class. The way we get the final grade is the reason why the class is hard. There is only one final three hour examination at the end of semester which determines your grade. It has six questions with multiple sub-questions. All sub-questions are connected in the sense that if you can't answer the first sub-question, you cannot answer any of the other remaining sub-questions. Most of the questions are complicated problems which took me about an hour to solve. I was required to finish six questions within three hours. I was slow and couldn't finish more than three in my first attempt. I failed the first time I took the examination. Almost more than half(Awk) the class failed the course and retook it. The next semester, I studied and practiced really hard. I passed the course but not with the grade I expected. One of my international friends could not pass the test the second time and he had to change his masters(P). In the end, I can say that taking the course was a great experience for me. I learned a lot.
8	Prompt 2: Name a controversial topic in your field and delineate your position.
9	Many physicists differ in their opinions on the existence of a unified theory which explains all (D) laws of nature. Termed as(X) the biggest problem of(PP in) physics for this century is the existence of a unified theory of everything which can explain all (D) phenomenon(S/Pl) of(PP in) nature. As of now, the laws of nature at minute (quantum) scales are described by quantum mechanics. The laws of nature at large (astronomical) scales are described using (D) general theory of relativity invented by Einstein. There are also some other theories and concepts (P/SPG) that are being investigated. But (PP in)
+ ≡ Error Tally Log ▾ Paragraphs ▾ Edit Log ▾ Symbols ▾	

Figure A2

Key to Codes

	A	B	C	D
1	Global Errors	VT	Verb tense	Verb tense must be changed
2		VF	Verb form	Verb form must be changed (aspect)
3		SS	Sentence structure	There is a sentence structure error (example: a run on sentence)
4		WO	Word Order	There is a word order error.
5		TW	Transition words	There is a problem with the connecting words (transitions)
6	Local Errors	SV	Subject-verb agreement	The subject and verb do not agree.
7		D	Determiner	There is a determiner error (a, an, the).
8		S/PL	Singular/plural	There is a singular/plural error.
9		C/NC	Count/non-Count	There is a count/non-count error.
10		WC	Word choice	The word does not fit well in this context. Choose another word.
11		WF	Word form	There is a word form error (it should be another part of speech)
12		PP	Preposition	There is an error with the preposition.
13	Other Errors	SPG	Spelling	There is a spelling error.
14		P	Punctuation	There is errored or missing punctuation.
15		C	Capitalization	There is a capitalization error – capitalize or lower case the word.
16		^	Missing a Word	There is a word missing.
17		X	Delete	A word needs to be erased.
18		?	Unclear Meaning	The meaning is unclear.
19		Awk	Awkward	There is a better way to say this
20				
21				
22				
23				

+ ☰ Paragraphs ▾ Symbols ▾ Error Tally Log ▾ Edit Log ▾

Figure A3

Example Error Log

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	
1																													
2		Prompt	1	2	3	4	5	Total	6	7	8	9	10	Total	11	12	13	14	15	Total	16	17	18	19	20	Total	21	22	
3		Paragraph Score																											
4		Global Errors	VT	1		1	4		6	2					2	1					1		1	1			2		
5			VF		1	1			2		2			1	3	3		2			5				1	1	2		
6			SS	1	1				2	1			1		2		1				1								
7			WO				1		1				1	1	2						1			1			1	1	
8			CW		1				1	1			1		2														
9			SV														2				2	1		1	1		3		
10			D	2	3	3	4	1	13				1	4	5	6	1	1	1	1	10		1	1	6	1	9		
11			s/pl		1	1	1		3				1		1	1	1	1			3	3			1		4		
12			c/nc																										
13			WC	1			3	1	5		1	2			3			1			1		1	3		4	8		
14			WF		1				1			2			2	1		1			2								
15		Local Errors	PP	1	3	2	2	2	10	1	1	1	1	3	7		1				1			1		1	2	1	
16			SPG	5			3		8	4	1			1	6	1	3				4		1		2	3	6		
17			P	7	1	4	7	1	19	5	4	5	4		18	1	1		1		3	6	1	2		4	13	4	
18			C			4			4	1			2		3			2		3	5								
19			^	1				2	3		2	3		2	7	2			1	1	4	2	2	1	1		6	2	
20			X	2	1	2		1	6		1		1	1	3							1	1		1		3	1	
21			?																										
22		Other Errors	Awk				1		1	1				1	2							1			1	1	3		
23			Total Errors	21	13	18	25	9	85	16	12	13	12	15	68	16	8	9	4	6	43	14	8	10	15	15	62	9	
24		Fluency	Word Count	244	164	153	252	119	932	125	163	153	151	151	743	151	106	117	145	125	644	161	154	172	225	169	881	142	132
25			# Sentences	18	9	9	17	9	62	9	9	9	10	11	48	9	7	6	7	7	36	9	10	9	13	15	56	9	10
26																													
27																													
28																													

Appendix C

Control Group Instructions

Figure C1

Paragraph Instructions Part 1

Paragraph Writing

You will be asked to write 3 paragraphs. Prompts will be provided for you.

You will spend 10 minutes on each paragraph (30 minutes total)

Please time yourself using a computer, clock, or phone and end as soon as 10 minutes are up.

These paragraphs are not meant to be part of an essay. They should stand alone.

Figure C2

Paragraph Instructions Part 2

Paragraph 1

Please write a single paragraph (multiple sentences) response to the following prompt.

PLEASE TIME YOURSELF WRITING. Write for 10 minutes.

The answer box will expand as you write.

Paragraph 1: Is there a superior pet? What is it and why? *

Your answer
