

Dynamic Written Corrective Feedback among Graduate Students: The Effects of Feedback Timing

Grant Eckstein, Maureen Sims, & Lisa Rohm

Dynamic written corrective feedback (DWCF) is a pedagogical approach that offers meaningful, manageable, constant, and timely corrective feedback on student writing (Hartshorn et al., 2010). It emphasizes indirect and comprehensive written error correction on short, daily writing assignments. Numerous studies have demonstrated that its use can lead to fewer language errors among undergraduate and pre-matriculated college writers (see Kurzer, 2018). However, the benefits of DWCF among second language (L2) graduate writers and the role of feedback timing have not been well examined. We analyzed timed writing samples over a 12-week intervention from 22 L2 graduate students who either received biweekly feedback on their writing throughout a semester, or postponed feedback until the last two weeks of the semester. Writing was analyzed for grammatical errors, lexical and syntactic complexity, and fluency. Results showed that neither timely nor postponed feedback led to significant improvement in grammatical accuracy or lexical complexity, but timely feedback did result in more fluent and complex writing. These findings suggest that the timing of feedback may be trivial for accuracy development but is more important for complexity among graduate writers. Teachers, teacher trainers, and writing administrators may use these insights as they plan curricula and design grammar and writing interventions.

La rétroaction corrective écrite dynamique (RCED) est une approche pédagogique qui propose une rétroaction significative, gérable, constante et opportune sur les rédactions des étudiants (Hartshorn et al. 2010). Elle insiste sur la correction complète et indirecte d'erreurs dans de courts devoirs de rédaction quotidiens. De nombreuses études ont démontré que son utilisation peut amener les rédacteurs de premier cycle ou pré-inscrits au collège à faire moins d'erreurs de langue (voir Kurzer, 2018). Cependant, les avantages de la RCED chez les rédacteurs diplômés de seconde langue (L2) et le rôle joué par l'opportunité de la rétroaction n'ont pas été bien étudiés. Nous avons analysé des échantillons de rédaction écrites en temps limité sur une période d'intervention de 12 semaines chez 22 étudiants diplômés de L2 qui recevaient de la rétroaction deux fois par semaine sur leurs rédactions pendant la durée du semestre, ou une rétroaction différée jusqu'à deux semaines avant la fin du semestre. Les rédactions ont été analysées pour découvrir les erreurs grammaticales, la complexité lexicale et syntaxique, ainsi que la fluidité. Les résultats ont montré que ni la rétroaction opportune, ni la rétraction différée ne se traduisaient par une amélioration marquée de la précision grammaticale ou

de la complexité lexicale, mais la rétroaction opportune menait à une rédaction plus fluide et plus complexe. Ces résultats suggèrent que l'opportunité de la rétroaction peut ne pas beaucoup influencer sur le développement de la précision, mais s'avère plus importante pour la complexité chez les rédacteurs diplômés. Les enseignants, les formateurs d'enseignants et les administrateurs de programmes de rédaction peuvent se servir de ces résultats lorsqu'ils planifient les programmes et conçoivent les interventions en grammaire et en rédaction.

Keywords: written corrective feedback, DWCF, graduate student, grammar, complexity

Written error correction is one of the most time-consuming activities employed by second language (L2) writing teachers (Polio, 2012) but is often thought to be essential in helping L2 students improve as writers. This is because linguistic control is central to effective written communication. Thus, teachers often engage in written corrective feedback (WCF) (which can be defined as any written comments or feedback on student writing) that is meant to improve linguistic accuracy (Ferris, 2006, 2011).

Researchers have studied the efficacy of WCF for years (see Kang & Han, 2015) and have found evidence of its usefulness in improving at least some elements of linguistic accuracy such as determiner errors (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2010a, 2010b; Ellis et al., 2008; Sheen, 2007; Sheen et al., 2009). This stands in contrast to arguments made by Truscott (1996, 1999, 2001, 2007; Truscott & Hsu, 2008) who considered WCF ineffective and atheoretical, and suggested that “grammar correction in second language (L2) writing classes should be abandoned” (1996, p. 327). However, as evidence of the benefits of WCF has mounted, practitioners have developed systems to implement manageable WCF in a unified learning approach (Ferris, 2004, 2006).

One particular approach to error correction is Dynamic Written Corrective Feedback (DWCF), which routinizes treatment of errors in learners' writing. DWCF is a direct application of skill acquisition theory which posits that knowledge develops from declarative to procedural to automatic (DeKeyser, 2007a, 2007b). When applied to linguistic accuracy, declarative knowledge reflects a learner's knowledge of a grammar rule. Procedural knowledge reflects an ability to rapidly apply that rule in writing, and automaticity is “faster execution, with less attention, and fewer errors” (Polio, 2012, p. 381). Practice and feedback are both foundational to this theory because it is assumed that practice transforms declarative knowledge into procedural and then automatic knowledge; feedback, in turn, is thought to prohibit the proceduralization of incorrect forms.

DWCF as Meaningful, Manageable, Constant, and Timely

In the DWCF approach, students produce short daily paragraphs, and later, with the aid of coded teacher feedback, edit the paragraphs over multiple rounds until they are error free. Teachers correct those paragraphs and provide further coded correction on subsequent drafts as needed. Because students are taught to interpret the codes, and they keep a record of their most frequent errors, the feedback is thought to be *meaningful* (Hartshorn et al., 2010). Further, the paragraphs should be short (timed at 10 minutes) in order to allow for *manageable* feedback, as the developers explain that teachers must have enough time to provide quality feedback, and students must have time to apply the feedback.

The DWCF approach is further characterized as constant and timely as in the following propositions adapted from Hartshorn et al. (2010).

Constant

1. Students produce new pieces of writing nearly every class period of the course.
2. Students receive feedback (student writing with coded symbols) nearly every class period of the course.

Timely

1. Student writing is consistently marked with coded symbols.
2. Student writing, with coded symbols, is consistently returned the following class period.

Some practitioners have found the constant process to be time-consuming and rigorous for students and teachers (see Eddington, 2014; McQuillan, 2012; Shelley, 2014), leading to the question of how important constant writing and feedback is as a feature of DWCF (e.g., Kurzer, 2018). In the present study, we go further and question the importance of timely feedback. While Hartshorn et al.'s (2010) initial description of these concepts makes them appear highly interdependent—that is, students should write regularly and receive regular feedback shortly after producing a piece of writing—these concepts can in fact operate separately (J. Hartshorn, personal communication, July 27, 2020). This results in four possible combinations of feedback timing as outlined in Table 1.

Table 1
Four DWCF Conditions of Constant and Timely Feedback

| | Timely | Non-timely |
|-----------|--|---|
| CONSTANT | Traditional DWCF approach See Hartshorn et al. (2010); Hartshorn & Evans (2015) | Constant paragraph writing with feedback postponed to the end of the semester Researched in our current study |
| IRREGULAR | Periodic or irregular paragraph writing but with timely feedback See Kurzer (2018) | Periodic or irregular paragraph writing with postponed feedback Not yet researched |

Although meaningfulness of feedback has been tied to sociocultural theory (i.e., Vygotsky, 1978; see also Kurzer, 2018) and manageability has been linked to cognitive load theory (Kirschner, 2002; Paas et al., 2004), there is less theoretical justification for timeliness of feedback—the idea that feedback should follow quickly after student performance. While timeliness is intuitively appealing and generally preferred by students, a possible theoretical justification for timely feedback is found in distributed practice research. This research posits that learning improves when practice sessions are spread out over time, or distributed, rather than massed (Rohrer & Pashler, 2007). Distributed practice, as it relates to feedback, entails multiple sequential feedback sessions over time rather than grouping all feedback into a condensed timeframe as with massed practice. Decades of research across multiple information types and skills have shown that students increase their long-term retention of studied material when time gaps separate practice sessions (for reviews, see Cepeda et al., 2006; Rohrer & Pashler, 2007).

When applying this concept to language development, researchers have hypothesized that students who regularly return to their writing after a timing gap in order to correct errors will develop long-term retention of grammatically accurate language. However, previous research on distributed practice in foreign language contexts has been inconclusive, contradictory, and belies expectation (e.g., Bird, 2010; Freed, Segalowitz, & Dewey, 2004; Serrano & Munoz, 2007). Freed et al. (2004), for instance, found that students in a massed-learning classroom environment (seven-week summer immersion program) outperformed those in a more distributed curriculum (12-week semester abroad program) in French oral fluency despite the fact that the students in both environments logged the same number of classroom hours. Serrano and Munoz (2007) found similar results when comparing three groups of English language learners. They all studied for 110 hours, but the instruction was spread over 5 weeks, 3–4 months, or 7 months. The latter (distributed) group was the weakest in posttest assessments of listening,

grammar, vocabulary, and reading. Bird (2010), on the other hand, compared 3-day and 14-day intervals of English grammar instruction in an ESL class and found that the longer distribution resulted in greater long-term retention of English syntax. These contradictory findings warrant further research into the distribution effect of language control, echoing DeKeyser's (2007b) call for more research to determine the amount and nature of feedback necessary during practice, especially in writing.

Furthermore, extant research on DWCF has only investigated effects of grammar correction on undergraduate and pre-matriculated university students. Little is known about whether DWCF is useful for international L2 graduate students, yet research suggests that, despite possessing relatively advanced language skills, graduate-level L2 writers still struggle to meet writing expectations in terms of grammar and writing conventions in English (Bitchener & Bastrukmen, 2006; Hyland & Milton, 1997; Kuzhabekova, 2020; Leki, 2010; Ma, 2019; Powers & Nelson, 1995). An analysis of written feedback received by international L2 graduate students studying in the United States showed that the highest frequency category was "language and writing" (Leki, 2010, p. 273). Such gaps in the knowledge and application of English writing conventions can be problematic for graduate-level writers, especially those aiming to publish (Kuzhabekova, 2020). Conversely, Kurzer (2018) speculates that some learners may be so advanced that error correction will have little effect. By extension, it is possible that graduate students demonstrate a ceiling effect, and therefore fail to benefit from DWCF in accordance with Truscott's argument against grammar correction. However, some evidence indicates that L2 graduate students, due to greater English proficiency, are able to improve accuracy while simultaneously focusing on other aspects of communication, such as pragmatics (Xu et al., 2009), suggesting that graduate students may be favourably positioned to gain from DWCF interventions.

This research study critically investigates the role of feedback timing by comparing one group of students who received timely WCF with a second group of similar students who received postponed feedback. The timely group wrote two 10-minute paragraphs per week (Tuesday and Thursday) throughout the semester and received WCF within two class periods (i.e., feedback was distributed throughout the semester). The postponed group also wrote two 10-minute paragraphs per week, but all feedback was postponed until the last two weeks of the semester (i.e., it was massed into a two-week time frame). In addition to exploring the effect of feedback timing on the student's writing accuracy, the study further investigates a previously overlooked learner background by examining the effects of DWCF on international L2 graduate student writers.

Review of DWCF Studies

Since 2010, there have been only six published investigations of DWCF on students' written English accuracy (Evans et al., 2010; Evans et al., 2011; Hartshorn et al., 2010; Hartshorn & Evans, 2012; Hartshorn & Evans, 2015; Kurzer 2018). Hartshorn et al. (2010) created a quasi-experimental design with intact classes of pre-matriculated students studying at an intensive English program (IEP). Participants included 19 control and 28 treatment students at different proficiency levels. Results of pre- and posttest timed writing over 15 weeks of daily DWCF showed a significant group difference with treatment students achieving significantly higher accuracy scores (error-free T-units/T-units) while control students actually decreased slightly in accuracy over time. The findings indicated that DWCF led to accuracy gains over a 15-week semester. A second study (Evans et al., 2010) demonstrated similar improvement in error-free clauses over a 13-week semester of daily DWCF among 22 IEP students, although no control group was used.

Evans et al. (2011) replicated the first study with undergraduate university students in a 15-week semester and likewise found that the treatment group ($n = 16$) significantly outperformed the control group ($n = 14$) in general written accuracy while the latter decreased slightly. When measured over 30 weeks among IEP students in a similarly designed follow-up study, the treatment group also significantly outperformed the control group (Hartshorn & Evans, 2015).

To measure which linguistic features responded to DWCF, Hartshorn and Evans (2012) evaluated written development in seven broad categories of sentences, verbs, numeric agreement, lexicon, determiners, semantic accuracy, and mechanics. Only the last three categories significantly distinguished control ($n = 19$) and treatment ($n = 28$) groups across a 15-week semester, suggesting that despite its overall benefit, DWCF may affect some language areas more than others. However, in a similar design with undergraduate students from three levels of English proficiency, Kurzer (2018) investigated differences in mechanical, local, and global errors over a 10-week quarter. Results showed that treatment groups ($n = 107$) significantly outperformed control groups ($n = 79$) on all measures.

These six studies represent consistent and positive evidence for the efficacy of DWCF in developing linguistic accuracy among both pre-matriculated and undergraduate students. On the other hand, these same studies demonstrate little evidence for improvements in written fluency and complexity, which may be at odds with accuracy. This is because writers might intentionally write less or use linguistically simplified structures in an attempt to improve linguistic accuracy, or as Hartshorn and Evans (2015) postulate, "gains in one aspect of writing come at the cost of another" (p. 14). Hartshorn et al. (2010) investigated measures of fluency and writing complexity where fluency was defined as number of words produced within the 30-minute time limit, and

complexity was a calculation of the average number of words per T-unit. Results from pre- and posttest essays indicated that neither of these measures demonstrated significant change for either the control or treatment group but that the treatment group showed slight, nonsignificant declines, causing the authors to speculate that the DWCF intervention “may have had a slight negative effect on writing fluency and complexity” (p. 100). Similar results were reported by Evans et al. (2011) among 30 undergraduate L2 writers. Hartshorn and Evans (2015) also found no difference between control ($n = 12$) and treatment ($n = 15$) groups in terms of fluency and syntactic complexity in an IEP setting, although one measure of lexical development (proportion of words from the second thousand most frequent words of the general service list) showed a decline among the treatment group. Thus, while DWCF appears to be an effective replacement for grammar instruction, it may need to be coupled with traditional process writing instruction to compensate for a potential lack of growth in both writing fluency and complexity.

In all studies described above, the methodology employed used timely feedback, i.e., delivered throughout the semester, which, despite being a defining tenet of DWCF and intuitively appealing as a distributed pedagogical intervention, is nonetheless laborious for students and teachers and tends to contrast with practical approaches to error correction which, when timed to teacher schedules, may not be timely at all. Furthermore, none of the six existing studies of DWCF have investigated student populations beyond the undergraduate level despite the potential for graduate students to not only benefit from this type of error correction but also to more effectively balance error correction with fluency and complexity development. To address these gaps in the literature, we have designed a study with graduate language learners utilizing DWCF with and without timely feedback. The following research question guided our inquiry: To what extent does *timely* DWCF, delivered throughout a semester of instruction, lead to changes in L2 speaking graduate students’ linguistic accuracy, lexical complexity, syntactic complexity, and fluency?

Methods

Context

This study collected data from three semesters of a research and writing course for international graduate students (master’s and PhD) taught as an elective in the linguistics department at a large research institution in the western United States. As a stand-alone course, students from any discipline and at any stage in their program were encouraged to enroll. The course met twice a week for 75 minutes over 15 weeks and was focused on improving students’ linguistic accuracy as well as writing literature reviews for theses,

dissertations, and research articles. Homework consisted primarily of multiple paragraphs assigned weekly that required students to practise various types of writing including intertextual summary, paraphrase, and critical analysis which culminated in the creation of a full literature review. Students also completed additional assignments such as completing assigned readings, providing peer reviews, responding to feedback from professors or peers in their department, and producing evidence-based research essays.

In addition, a signature component of this course was the writing and revising of biweekly timed (10-minute) in-class paragraphs focused on error correction using the DWCF approach (see Hartshorn et al., 2010 for full description of this method). The process we used is described in greater detail below, but in brief, the approach requires students to write a paragraph on a familiar topic, such as “Why did you choose your major?” or “What is a controversy in your field?” in class with a set time limit of 10 minutes. The instructor reviews these paragraphs after class and marks each one comprehensively for errors using codes for each error type. Students then correct the errors outside of class and resubmit the paragraph for a second round of coding by the instructor followed by further error correction and submission of a final third draft. Some of the 10-minute paragraphs were analyzed as part of this study.

The use of corrective feedback in the class was encouraged by the culture of the university, the linguistics department, and the objectives of the course. At the university level, the writing center provides grammar support to graduate and undergraduate L1 and L2 students and was led at the time of this study by directors who acknowledge the value of language control for effective communication and who regularly interact with L2 writing specialists across the campus and nationally. The composition department likewise encourages teachers to offer language instruction in first-year composition. DWCF in particular was originated by faculty in the linguistics department and is used widely in the school’s intensive English program as well as in first-year composition courses for international students. Its use in the present course for graduate L2 writers was encouraged by course objectives and was valued by the instructor who has employed it regularly in his instruction. Students in the present course and in other classes where DWCF is used regularly have reported their deep appreciation for DWCF on student evaluations.

Participants

Twenty-three international graduate students, nine women and 14 men, ranging from 24 to 39 years old ($M = 28$) participated in this research (see Table 2). The participants represented 10 language backgrounds and 14 academic disciplines and could be characterized as highly motivated as all were working towards theses, dissertations, or other research projects; several

had already published academic papers in English and/or other languages. All 23 participants were enrolled in one of three semesters of an intact writing and research course at the university. This stand-alone course was taught during three separate semesters, and students could take the course at any point in their education, thus semester assignment did not reflect or affect proficiency. All students were assigned as members of either a *timely* or *postponed* feedback group. Two of the three semesters (Winter & Fall of the same year) comprised the *timely* group, whereas the remaining semester (Fall of the previous year) constituted the *postponed* feedback group.

Table 2
Demographic Information of Study Participants by Group

| | Timely Feedback Group | Postponed Feedback Group |
|------------------|--|---|
| Number | 12 | 11 |
| Age (average) | 28 | 28 |
| Sex | 5 female, 7 male | 4 female, 7 male |
| Native Languages | Chinese, Portuguese, Haitian Creole, Korean, Hindi, Bengali, Spanish | Turkish, Spanish, Korean, Hindi, Telugu, Persian, Chinese |
| Program of Study | Second Language Teaching, Public Administration, Public Health, Mass Communications, Civil Engineering, Chemistry, Business Administration, Physiology and Developmental Biology, Media, and Performance Studies | TESOL, Chemistry, Physical Medicine and Rehabilitation, Physics, Exercise Sciences, Wildlife and Wildlands Conservation, Civil Engineering, Physiology, and Developmental Biology |
| Graduate level | 7 master, 5 doctorate | 5 master, 6 doctorate |

All semesters were taught by the first author who, at the time of the study, was the ESL writing program administrator in the Linguistics department and had more than 10 years of experience using DWCF and teaching L2 writing. He was assisted by the third author, a TESOL master's student teaching assistant (TA) with two years of experience working as a TA for graduate courses and four semesters of experience teaching ESL writing using DWCF at an IEP. The research for this study formed part of the third author's master's thesis; all work was supervised and reviewed by the course instructor, the first author.

Materials

Throughout the semester, the students were required to respond in writing to a total of 20–22 personal prompts (see Appendix A). The writing took place in class and was preceded by a short group discussion of the prompt to stimulate thinking. Each time, the students were given 10 minutes to respond to a prompt and were encouraged to focus on grammatical accuracy in their responses. All paragraphs were written and corrected using individual Google Sheets, which could be viewed at any time by the student and teacher and did not include spellcheck or suggestions for grammar improvement. One tab contained the DWCF feedback codes and their meanings, and another was used for the error log where students recorded the number of each error type identified by the TA on their first draft; this log allowed students to track their most common errors across paragraphs and focus on specific areas of weakness. An example of these documents is found in Appendix B.

The feedback codes employed for this study included traditional DWCF error codes (Hartshorn et al., 2010; see Appendix C for full list) which were then grouped into the three categories of global, local, and “other” errors similar to the process used by Kurzer (2018; see also Lane & Lange, 2012). As shown in Appendix C, global errors included verb tense, sentence structure, and word order issues; local errors included subject-verb agreement, determiner and singular/plural issues; other errors included what Kurzer termed “mechanical errors” such as spelling, punctuation, and awkward wording.

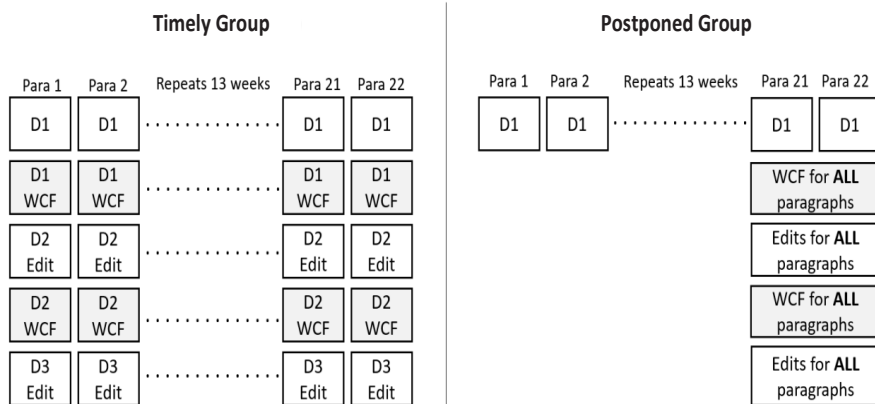
Procedures

Our study followed a quasi-experimental design with two intact classes as treatment groups receiving 15 weeks of DWCF intervention as part of the aforementioned writing course. Students were divided into timely feedback and postponed feedback groups based on the semester in which they took the course. Both groups wrote two paragraphs weekly and produced the same amount of writing, a total of 22 paragraphs over the semester (see Figure 1). The timely feedback group received error feedback within two class periods of submitting a draft while the postponed feedback group received feedback on all their paragraphs during the last two weeks of the semester (i.e., it was postponed until the end of the semester). Once feedback was received, both groups were required to edit their paragraphs and resubmit them for a second round of feedback. The timely group received and processed timely corrective feedback throughout the semester while the postponed group received and processed corrective feedback on 22 paragraphs during the last two weeks of the semester, which is essentially the equivalent of editing a substantial term paper near the end of a course and is not particularly uncommon for graduate students (Onwuegbuzie & Collins, 2001). After receiving feedback

on their second drafts, students were required to again edit their paragraphs and submit them a final time.

Errors were marked with indirect, coded feedback in the first draft. Errors which remained or were newly introduced in a subsequent draft were marked only. Coded and direct feedback was provided for preposition errors for which explicit grammar rules are lacking (i.e., they are less treatable errors; see Ferris, 1999) and otherwise often lead to guessing. When mechanical or word choice errors proved especially subtle, such as the need to include a hyphen in an otherwise correctly spelled word, direct feedback was occasionally paired with an error code. Besides these instances, direct coding was limited, and the students' final drafts were left unmarked. Metalinguistic explanations of errors were exceptionally rare but occurred when a student asked for specific feedback. Our study differs from prior DWCF research in that we did not require students to continue editing until their paragraphs were completely error free. Instead, students were expected to write a total of three drafts, which for many resulted in error-free writing, since the DWCF component was only a portion of the larger writing curriculum (see Figure 1).

Figure 1
DWCF Paragraph and Editing Progression for Both Groups



*Both groups completed Draft 1 (D1) in class; all others were completed outside of class

Data Analysis

Student writing was collected for analysis at the beginning and end of the semester using the first draft of the first three 10-minute paragraphs and the first draft of the last three 10-minute paragraphs. The first draft of all 6

paragraphs was used as it provided an unedited sample and thus represented the level of writing accuracy of each student. The first draft of the first three paragraphs was combined to provide the *pre* data of student writing accuracy, whereas the first draft of the last three paragraphs was combined to provide the *post* data of student writing accuracy. Three paragraphs for each, *pre* and *post*, were chosen in order to establish a more consistent baseline of accuracy scoring for each student and thus minimize the impact of potential outlying data if students had, for example, a “bad” day. Further, using timed, in-class 10-minute paragraphs allowed researchers to collect the equivalent of 30 minutes of controlled writing sampled from a variety of topics while remaining faithful to the typical class design and assignment structure (outlined above) that has been established for this course. Accuracy data were calculated by tabulating the number and type of errors produced in each text using traditional DWCF coding described above (see Hartshorn et al., 2010). All paragraphs were coded by a member of the research team who had completed training and norming sessions prior to beginning the experiment.

Like Kurzer (2018), we normalized accuracy data by dividing the number of errors in a category by the number of words in the text and multiplying this number by a constant. While Kurzer used 100 as the constant, we used 481, which is the average number of words for all essays in the sample and follows common normalizing practices (see Biber et al., 1998; Eckstein & Ferris, 2018; Ferris, 2006).

To quantify the data for further analysis, lexical, syntactic, and fluency data were all calculated automatically using the web-based interface of lexical complexity and L2 syntactic complexity analyzers (Ai & Lu, 2013; Lu, 2010, 2011; Lu & Ai, 2015). Lexical complexity was reflected by measures of lexical density, which is the ratio of lexical (not grammatical) words to total words as in $\text{tokens}_{\text{lexical}}/\text{tokens}_{\text{total}}$ (Lu, 2010; Ure, 1971). Lexical sophistication was calculated based on sophisticated words, defined by Lu (2010) as words “not on the list of the 2,000 most frequent words generated from the British National Corpus” (p. 192); the lexical sophistication formula was therefore $\text{type}_{\text{soph}}/\text{type}_{\text{total}}$. The third measure was type-token ratio. These three measures were selected to 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.

Syntactic complexity was reflected by measures of mean sentence length, mean length of T-unit, and complex nominals per clause. Lu (2011) found that mean sentence length and complex nominals per clause (e.g., adjective + noun, possessives, prepositional phrases) were strong discriminators of English essay quality among Chinese writers. Lu and Ai (2015) found that all three measures reflected significant differences in argumentative essays written by non-native English students. Further, nominal clauses, including

noun clause modification, have been related to essay quality (Taguchi et al., 2013) and clausal compression in academic writing (Biber & Gray, 2010).

For fluency measures, we used the number of words, sentences, and clauses produced within three timed paragraphs, amounting to 30 minutes of writing. Other DWCF studies have likewise used raw word count as a fluency measure (e.g., Hartshorn et al., 2010). We additionally measured sentence and clause counts since clausal complexity can indicate relative formality of writing (i.e., oral or written register; see Biber et al., 2011; Taguchi et al., 2013).

Accuracy, lexical and syntactic complexity, and fluency data from the first draft of the first and last three 10-minute paragraphs were coded as dependent variables and analyzed using two-way ANOVAs (group x time). Jamovi version 1.0.7.0, a free software analysis system, was used to run the statistical analyses (Fox & Weisberg, 2018; R Core Team, 2018; The jamovi project, 2019).

Results

Accuracy

As illustrated in Table 3, the three categories of grammatical accuracy showed a small reduction in errors from pre- to posttest error scores for the timely feedback group. Conversely, the postponed feedback group increased in all error categories over time. However, these differences were not statistically significant for either time or treatment (all $ps > .5$; all $\eta^2_p < .008$).

Table 3
Accuracy Scores across Groups and Tests

| Error type | Group | <i>M(SD)</i> | | | <i>MD</i> |
|------------|-----------|--------------|------------|------------|------------|
| | | <i>n</i> | Pre | Post | Difference |
| Global | Timely | 12 | 8.38(6.95) | 6.15(5.12) | 2.23 |
| | Postponed | 11 | 7.35(3.71) | 8.51(4.01) | -1.16 |
| Local | Timely | 12 | 24.6(15.7) | 20.5(9.32) | 4.1 |
| | Postponed | 11 | 19(10.7) | 23.2(15.8) | -4.2 |
| Other | Timely | 12 | 26.1(12.2) | 21.9(12.4) | 4.2 |
| | Postponed | 11 | 23.1(12.5) | 28.8(12.2) | -5.7 |

Lexical Complexity

Descriptive measures of lexical complexity showed that it generally decreased between pretest and posttest writing for both treatment groups. The only

exception was that of type-token ratio which increased across time for the timely feedback group (see Table 4). Results of a two-way ANOVA (group x time) showed no significant difference for treatment type (all $ps > .15$; all $\eta^2_p < .05$). However, test time was significant, showing that lexical density decreased over time ($F(1, 22) = 5.42, p = .025, \eta^2_p = .11$) from a mean of .51 to .49 as did lexical sophistication (number of sophisticated types/total types) ($F(1, 22) = 4.83, p = .034, \eta^2_p = .10$) from a mean of .24 to .21.

Table 4
Lexical Complexity across Groups and Tests

| | | <i>M (SD)</i> | | | <i>MD</i> |
|------------------------|-----------|---------------|--------------|--------------|------------|
| | | <i>n</i> | Pre | Post | Difference |
| Lexical Density | Timely | 12 | 0.519(0.033) | 0.510(0.038) | 0.009 |
| | Postponed | 11 | 0.518(0.029) | 0.484(0.025) | 0.034 |
| Lexical Sophistication | Timely | 12 | 0.238(0.058) | 0.223(0.057) | 0.015 |
| | Postponed | 11 | 0.234(0.034) | 0.188(0.025) | 0.046 |
| Type-Token Ratio | Timely | 12 | 0.427(0.070) | 0.447(0.061) | -0.02 |
| | Postponed | 11 | 0.470(0.107) | 0.439(0.065) | 0.031 |

Syntactic Complexity

Results for syntactic complexity show a consistent pattern of reduction over time for both treatment groups as shown in Table 5. However, a two-way ANOVA (group x time) showed that the main effect of time was not significant for any of the measures. The main effect comparing treatment groups, on the other hand, was significant for all measures (all $ps \leq .001$) showing that those receiving postponed feedback wrote about double the sentence size and mean length of T-units but produced about half as many complex nominal clauses as those receiving timely feedback (see Table 6).

Table 5
Syntactic Complexity across Groups and Tests

| | | <i>M(SD)</i> | | | <i>MD</i> |
|-----------------------------|-----------|--------------|------------|------------|------------|
| | | <i>n</i> | Pre | Post | Difference |
| Mean Sentence Length | Timely | 12 | 9.23(9.53) | 7.32(7.02) | 1.91 |
| | Postponed | 11 | 16.2(5.53) | 16.1(1.88) | 0.1 |
| Mean Length of T-unit | Timely | 12 | 6.84(7.03) | 6.2(6.47) | -6.96 |
| | Postponed | 11 | 14.4(4.77) | 13.8(2.34) | 12.96 |
| Complex Nominals per Clause | Timely | 12 | 1.48(0.57) | 1.44(0.74) | 0.04 |
| | Postponed | 11 | 1.15(0.28) | 0.72(0.13) | 0.43 |

Table 6
Significant Main Effect of Treatment Group for Syntactic Complexity

| Group | | <i>n</i> | <i>M</i> | <i>SD</i> | <i>F</i> | <i>p</i> | η^2_p |
|------------------------|-----------|----------|----------|-----------|----------|----------|------------|
| Mean Sentence Length | Timely | 24 | 8.27 | 8.24 | 15.87 | <.001 | 0.27 |
| | Postponed | 22 | 16.10 | 4.03 | | | |
| Mean Length of T-unit | Timely | 24 | 6.52 | 6.62 | 21.60 | <.001 | 0.34 |
| | Postponed | 22 | 14.10 | 3.68 | | | |
| Complex Nominal Clause | Timely | 24 | 1.46 | 0.65 | 12.48 | .001 | 0.23 |
| | Postponed | 22 | 0.93 | 0.31 | | | |

Fluency

Measures of fluency showed increases across the three measures except for the postponed feedback group whose word count decreased from the pretest to the posttest (see Table 7). The measure of word count resulted in the only statically significant main effect, suggesting that the participants who received timely feedback ($M = 530$) outperformed those who received postponed feedback ($M = 428$) $F(1, 22) = 5.15, p = .03, \eta^2_p = .11$.

Table 7
Fluency across Groups and Tests

| | | <i>M(SD)</i> | | | <i>MD</i> |
|---------------------|-----------|--------------|------------|------------|------------|
| | | <i>n</i> | Pre | Post | Difference |
| Number of Words | Timely | 12 | 523(118) | 536(145) | -13 |
| | Postponed | 11 | 441(166) | 414(176) | 27 |
| Number of Sentences | Timely | 12 | 27.7(6.23) | 32.3(11.1) | -4.6 |
| | Postponed | 11 | 25.3(8.93) | 25.5(10.3) | -0.2 |
| Clauses | Timely | 12 | 59.8(15.4) | 65.1(21.3) | -5.3 |
| | Postponed | 11 | 51(21.3) | 54.9(25.4) | -3.9 |

Discussion

The purpose of this study was to determine whether timely feedback is needed for language development using the DWCF approach for error correction, and the results offer some rather salient insights. Firstly, students' error scores reduced over the course of traditional (timely) DWCF treatment as has been shown in other studies (see Kurzer, 2018). In previous DWCF studies this reduction has been statistically significant, whereas in this study, the gains were not significant. Two features of this study may have played a role. First of all, the major novelty in this study is that participants were graduate students suggesting that these writers may reach a ceiling in their written linguistic accuracy unlike pre-matriculated and undergraduate writers as Kurzer (2018) speculated. Secondly, the prompts used in this study were also focused on students' personal interests (e.g., "Why did you choose your area of study?") which may have elicited language students were already familiar with and which they may have had linguistic resources to address.

Further results showed that while timely feedback seemed to make a difference in terms of reducing linguistic errors, postponed feedback actually increased those errors, though this may also suggest that learners became more aware of issues in their writing only at the end of the semester. However, inferential tests showed that the linguistic errors between the two groups were not significantly different. Nominally, this indicates that neither treatment had a significant effect on errors over 13 weeks. This brings into question whether DWCF actually needs to include timely feedback or if merely writing regularly with the eventual expectation of error correction is sufficient to bolster students' linguistic accuracy. The latter option may be more appealing to teachers who have limited time to provide written feedback. However, given that the descriptive data show a positive trend for

timely feedback and a negative one for postponed feedback, we speculate that a longer treatment period could lead to greater divergence between the two groups, potentially resulting in a statistically significant difference between them in terms of linguistic accuracy.

The lexical complexity data supported findings that hitherto have been supported by descriptive analysis only. While Hartshorn and Evans (2015) found a decline in students' use of the second thousand most frequent words from the general service list, this study found significant reductions in all participants' lexical density and lexical sophistication over time, suggesting that DWCF may negatively affect writers' lexical development regardless of timing of feedback, likely because writers increasingly relied on a core set of vocabulary in an effort to reduce lexical errors. One exception to this trend was type-token ratio in which the postponed feedback group showed greater lexical variety while the timely feedback group showed less. While the result was nonsignificant, this finding suggests that more timely error correction may have partially affected the change. To explore this, a post-hoc textual review of students' original writing was performed, and it suggested that the diversity within the postponed group was caused more by spelling errors than by novel vocabulary usage.

The syntactic complexity results likewise confirm what has only been speculated upon in previous research. While several studies have observed nonsignificant declines in syntactic complexity over treatments of DWCF (i.e., Evans et al., 2011; Hartshorn & Evans, 2015; Hartshorn et al., 2010), the present study observed significant differences showing that postponed feedback recipients wrote more words per sentence and T-unit than timely feedback recipients, but they produced significantly fewer complex nominals. Taken together, we interpret this to mean that graduate student writers in this study produced more compact and complex writing when feedback was timely. This is an important observation because academic writing has been shown to rely on more phrasal complexity than mere clausal complexity (Biber & Gray, 2010; Biber et al., 2011). We speculate that timely feedback gave students greater confidence to use structural compression techniques (Biber & Gray, 2010), such as adjectival pre-modification (e.g., "*expected findings*"), noun pre-modification (e.g., "*color photo*") and prepositional phrase post-modification (e.g., "*They examined the effects of arsenic on solubility*"), which were discussed in course instruction as ways to make writing more concise.

The fluency data in this study partially contradicted findings reported elsewhere that fluency declined across DWCF interventions (Evans et al., 2011; Hartshorn et al., 2010). Hartshorn and colleagues have reported nonsignificant declines in fluency measures, but this study showed increases in sentences and clauses produced over time, although these gains were also nonsignificant. The number of words demonstrated the only significant finding within the fluency data: the timely feedback group wrote significantly more words than the postponed feedback group. It is possible that timely

feedback led writers to be more sure of themselves as has been reported in the literature (see Ruegg, 2018; Warden, 2000); whereas, the postponed feedback group may have felt a lack of such motivation. Graduate writing research is rife with evidence showing that a lack of feedback opportunities can negatively affect writing motivation and success (see Kamler, 2008; Mannon, 2016).

Ultimately, the main goal of this study was to investigate the role of timely error correction in DWCF among graduate students. While timely feedback appeared to have no significant impact on linguistic accuracy, it did support syntactic complexity and fluency development. This observation lends credence to the use of DWCF as a language tool and offers at least an initial glimpse into its value as a tool for academic writing development as well.

Conclusion

When taking all data from this study together, it seems that feedback timing did not substantially affect error correction, but it did have an effect on complexity and fluency. Writers who received postponed feedback wrote longer but fewer sentences while those receiving timely feedback wrote fewer, more complex sentences. Previous research suggested that DWCF treatment may lead to impoverished fluency and complexity—and the present data partially support that—but we contend that a reduction in fluency or complexity may be unproblematic and even desirable among graduate writers who are obligated to write succinctly and adopt complex phrasal structures when socializing into academic writing. Given this interpretation, it appears that timely feedback is more desirable than postponed feedback for graduate students because it overlaps with compact, complex writing.

These observations have implications for graduate language teachers, teacher trainers, and administrators. One subtle implication is that timing of feedback may be more trivial than previously expected when focusing on linguistic accuracy—potentially because of a ceiling effect among graduate writers. Teachers who find it difficult to produce timely feedback may take some comfort in this, and teacher trainers can benefit from this insight when offering instruction on feedback timing. However, when accuracy alone is not the aim but rather complexity in writing, timely feedback is useful, and so DWCF may play a larger role in writing development among graduate writers than previously reported for undergraduates. We therefore recommend the use of DWCF as a supplement in graduate writing courses.

The role of context in these findings is important given the increasing push for internationalization within higher education, particularly in the United States and Canada (Anderson, 2015; Cheng et al., 2004). Thus, the universality of graduate-level linguistic support for international writers can transcend national boundaries, especially as the United States and Canada similarly grapple with increasing numbers of international graduate

students on their campuses. Anderson (2015) specifically suggests that even while Canadian universities do provide some options to help international L2 students succeed culturally and linguistically, the government's efforts to increase international enrollments means more students will need "more comprehensive and targeted academic support" (p. 179).

Further, this study examined the combination of graduate writing and corrective feedback and was situated in a classroom setting designed specifically to help graduate students improve their academic writing over a single semester. While this context benefitted students' fluency and syntactic complexity, alternative contexts may provide more or less support. The traditional advisor-advisee relationship may allow for more personalized language feedback and instruction (Dong, 1998), but time and resource constraints can also complicate timely corrective feedback (Kamler, 2008; Mannon, 2016). Additionally, universities may lack a dedicated L2 graduate research writing course (Kuzhabekova, 2020; Simpson, 2012), and therefore students may need to seek feedback from writing center tutors, peers, or paid editors to help them develop academic writing skills (Conrad, 2019; Eckstein et al., 2017).

Of course, the findings from this study are limited in generalizability and interpretability because of the small sample size and the short duration of the study localized at a single university where data were collected from multiple intact classes over three subsequent semesters. Additionally, a control or contrast group who received no or traditional grammar intervention was not included in this study. This omission prevents an objective evaluation of DWCF or graduate student writing that should be investigated in future research. Future studies should include more participants and evaluate a longer treatment time in order to further investigate what we speculate is a ceiling effect among graduate writers in terms of linguistic accuracy. Research of this nature should also tie student performance to their precise language proficiency levels since proficiency may play an important role in predicting accuracy outcomes. Additionally, the complex interplay of constant and timely as tenets of DWCF may need further exploration given that there are four possible combinations of these terms. Also, because constant DWCF refers both to writing paragraphs constantly and receiving constant feedback, more research is yet needed to help decipher the various processes that the four combinations may refer to.

While this study included some elements of ecological validity by placing error correction within a wider writing curriculum, limiting rounds of error correction, and offering postponed feedback, all of which are more likely to reflect authentic writing conditions, future research might go further. This could include examining longer pieces of writing and possibly multi-draft or take-home essays so that a greater amount of writing can be analyzed rather than the average of 481 words at the beginning and end of the semester produced by participants in this study. While it is possible that students

could enlist help from peers, tutors, and others and thereby demonstrate inflated writing ability, if writers were asked to report the assistance they received or were limited in their access to external support, such research could demonstrate how accuracy and writing develop in less stressful and more naturalistic environments. Students could also be asked about their impressions of DWCF, the timing of feedback, and their writing development as a way to understand the multidimensional impacts of DWCF.

The Authors

Grant Eckstein is a professor of linguistics at Brigham Young University where he teaches graduate academic writing and teacher training courses. His research interests include second language reading and writing development and pedagogy. He is the associate editor of the *Journal of Response to Writing*.

Maureen Sims is an independent researcher with an MA in TESOL from Brigham Young University. She teaches English language classes and researches in the areas of language assessment and L2 writing. She has published articles in journals such as *The Reading Matrix* and *Educational Measurement: Issues and Practice*.

Lisa Rohm is a TESOL MA graduate student at Brigham Young University. In addition to researching issues of grammar and writing pedagogy, she teaches these subjects at the university's English Language Center.

References

- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). John Benjamins. <http://doi.org/dxsd>
- Anderson, T. (2015). Seeking internationalization: The state of Canadian higher education. *Canadian Journal of Higher Education*, 45(4), 166–187.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press. <http://doi.org/dxsc>
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. <http://doi.org/ff6fvt>
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35.
- Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, 31, 635–650. <https://doi.org/10.1017/S0142716410000172>
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118. <http://doi.org/b8p6qx>
- Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes*, 5(1), 4–18. <http://doi.org/frnhsf>
- Bitchener, J., & Knoch, U. (2008). The value of a focused approach to written corrective feedback. *ELT Journal*, 63(3), 204–211. <https://doi.org/10.1093/elt/ccn043>
- Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten-month investigation. *Applied Linguistics*, 31(2), 193–214. <http://doi.org/dng6fg>
- Bitchener, J., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207–217. <http://doi.org/c62n2c>

- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380.
- Cheng, L., Myles, J., & Curtis, A. (2004). Targeting language support for non-native English-speaking graduate students at a Canadian university. *TESL Canada Journal*, 21(2), 50–71. <http://doi.org/dxr9>
- Conrad, N. (2019). Revisiting proofreading in higher education: Toward an institutional response to editors Canada's guidelines for ethical editing of student texts. *TESL Canada Journal*, 36(1), 172–183. <https://doi.org/10.18806/tesl.v36i1.1309>
- DeKeyser, R. (Ed.). (2007a). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press. <http://doi.org/cfv7bw>
- DeKeyser, R. (2007b). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–114). Erlbaum.
- Dong, Y. R. (1998). Non-native graduate students' thesis/dissertation writing in science: Self-reports by students and their advisors from two U.S. institutions. *English for Specific Purposes*, 17(4), 369–390. [https://doi.org/10.1016/S0889-4906\(97\)00054-9](https://doi.org/10.1016/S0889-4906(97)00054-9)
- Eckstein, G., Evans, K., Moglen, D., & Whitener, W. (2017). Graduate writing groups: An interdisciplinary approach to writing productivity. *The Journal of Teaching English for Specific and Academic Purposes*, 5(1), 1–11. <https://doi.org/10.22190/JTESAP1701001E>
- Eckstein, G., & Ferris, D. R. (2018). Comparing L1 and L2 texts and writers in first-year composition. *TESOL Quarterly*, 52(1), 137–162. <https://doi.org/10.1002/tesq.376>
- Eddington, B. E. (2014). *A modified approach to the implementation of dynamic written corrective feedback* [Master's thesis, Brigham Young University]. Retrieved from All Theses and Dissertations database. (Paper 4389). <http://hdl.lib.byu.edu/1877/etd6832>
- Ellis, R., Sheen, Y., Murakami, M., & Takshima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System* 36(3), 353–371. <http://doi.org/bjtzrd>
- Evans, N. W., Hartshorn, K. J., McCollum, R., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research*, 14(4), 445–463. <http://doi.org/d4b6qr>
- Evans, N. W., Hartshorn, K. J., & Strong-Krause, D. (2011). The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System*, 39(2), 229–239. <http://doi.org/d9krm9>
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(1), 1–11. <http://doi.org/ff6zpd>
- Ferris, D. (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. <http://doi.org/chrcwd>
- Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge University Press. <http://doi.org/dxr8>
- Ferris, D. R. (2011). *Treatment of error in second language student writing*. University of Michigan Press. <http://doi.org/dxttd>
- Fox, J., & Weisberg, S. (2018). *car: Companion to Applied Regression* [R package]. <https://cran.r-project.org/package=car>.
- Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26(2), 275–301. <http://doi.org/dx5f9f>
- 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. (2015). The effects of dynamic written corrective feedback: A 30 week study. *Journal of Response to Writing*, 1(2), 6–34.

- Hartshorn, K. J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, 44(1), 84–109.
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183–205. <http://doi.org/bd6gmx>
- Kamler, B. (2008). Rethinking doctoral publication practices: Writing from and beyond the thesis. *Studies in Higher Education*, 33(3), 283–294. <https://doi.org/10.1080/03075070802049236>
- 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. <http://doi.org/dxvm>
- Kirschner, P. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction*, 12(1), 1–10. <http://doi.org/d8vv8w>
- Kurzer, K. (2018). Dynamic written corrective feedback in developmental ESL writing classes. *TESOL Quarterly*, 52(1), 5–33. <http://doi.org/dxvn>
- Kuzhabekova, A. (2020). "There are notable linguistic problems": Publishing as a non-native speaker of English. In T. Ruecker & V. Svihla (Eds.), *Navigating challenges in qualitative educational research* (pp. 193–205). Routledge. <http://doi.org/dzj6>
- Lane, J., & Lange, E. (2012). *Writing clearly: Grammar for editing* (3rd ed.). Heinle Cengage Learning.
- Leki, I. (2010). "You cannot ignore": L2 graduate students' response to discipline-based written feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in 2nd language writing: Contexts and issues* (pp. 266–285). Cambridge University Press.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <http://doi.org/fj36jz>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writer's language development. *TESOL Quarterly*, 45(1), 36–62.
- 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. <http://doi.org/f7rvjp>
- Ma, L. P. F. (2019). Academic writing support through individual consultations: EAL doctoral student experiences and evaluation. *Journal of Second Language Writing*, 43, 72–79. <http://doi.org/dzj7>
- Mannon, B. O. (2016). What do graduate student want from the writing center? Tutoring practices to support dissertations and thesis writers. *Praxis: A Writing Center Journal*, 13(2), 59–64.
- McQuillan, J. (2012, January 20). *All correction, all the time: Is written error correction worth the effort?* [Blog post]. Retrieved from <http://backseatlinguist.com/blog/?p=39>
- Onwuegbuzie, A. J., & Collins, K. M. T. (2001). Writing apprehension and academic procrastination among graduate students. *Perceptual and Motor Skills*, 92(2), 560–562.
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32, 1–8. <http://doi.org/dp3ggm>
- Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, 21(4), 375–389. <http://doi.org/f4jt63>
- Powers, J., & Nelson, J. (1995). L2 writers and the writing center: A national survey of writing center conferencing at graduate institutions. *Journal of Second Language Writing*, 4(2), 113–138. <http://doi.org/dfbrrq>
- R Core Team. (2018). *R: A Language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, 16(4), 183–186. <http://doi.org/ckrgwd>
- Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *The Language Learning Journal*, 46(2), 87–102.
- Serrano, R., & Munoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System*, 35, 305–321.

- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255–283. <http://doi.org/dxvp>
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37(4), 556–569. <http://doi.org/fmsw6>
- Shelley, A. (2014). *Dynamic written corrective feedback: Achieving manageability* [Master's thesis, Brigham Young University]. Retrieved from All Theses and Dissertations database. (Paper 4109)
- Simpson, S. (2012). The problem of graduate-level writing support: Building a cross-campus graduate writing initiative. *Writing Program Administration*, 36(1), 95–118.
- Taguchi, N., Crawford, B., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, 47(2), 420–430. <http://doi.org/gf3bsd>
- The jamovi project. (2019). *jamovi*. (Version 1.0) [Computer Software]. <https://www.jamovi.org>.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369. <http://doi.org/bd55rq>
- 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, 111–122. <http://doi.org/dsdrv>
- Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics*, 27(2), 93–108.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272. <http://doi.org/b8f8dd>
- Truscott, J., & Hsu, A. Y-P. (2008). Error correction, revision, and learning. *Journals of Second Language Writing*, 17(4), 292–305. <http://doi.org/d959rq>
- Ure, J. (1971). Lexical density: A computational technique and some findings. In M. Coulter (Ed.), *Talking about text* (pp. 27–48). University of Birmingham.
- Vygotsky, L. (1978). Interaction between learning and development. In M. Gauvain & M. Cole (Eds.), *Readings on the Development of Children* (2nd ed, pp. 29–36). W.H. Freeman and Company.
- Warden, C. A. (2000). EFL business writing behaviors in differing feedback environments. *Language Learning*, 50(4), 573–616.
- Xu, W., Case, R., & Wang, Y. (2009). Pragmatic and grammatical competence, length of residence, and overall L2 proficiency. *System: An International Journal of Educational Technology and Applied Linguistics*, 37(2), 205–216.

Appendix A: Paragraph 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 [current university]? |
| 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: Student Example

| Paragraph Revisions | | Comments |
|---|--|---------------------------------------|
| <p>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.</p> | | |
| Other prompt: | | |
| Prompt 1: | Describe the most difficult course you have ever taken. | |
| Draft | <p>The most difficult course I have ever taken is(PT) English 1 (P)(P) which I took when I was an undergraduate student in [country]. It was difficult because I was terrible in(PP at) English. Before I went to a(D) university, I studied English only in(PP X) reading and listening(WO). (D) Topics of (*D) materials I had learned were extremely difficult and boring. It was because I had learned English just for (D) [country's] SAT ^ to go to university. I was a student who studied (native language) in sneak(WC) in English class, since I hated English then. When(X) I took the "(P)English 1(P)" class was(X) right before I entered in(X) my university. I lost the goal to keep learning English then(? I'm not sure what you mean by this), and I wanted to avoid the class(SP) but it was mandatory. It was hard for me to get used to ^ in the(D) environment that everyone speak(VT) and listen(VT) (*PP in) English. The worst part of the class was (D) short presentation. Even though I had a chance to pick topic which I like(VT), I kept postponed(VT/VF) to prepare(VF) it. Finally(P) on the presentation day, I felt so nervous and wanted to run away(P)</p> | (PP X) or (PP x) means no preposition |
| Rewrite | <p>The most difficult course I have ever taken was English 1, which I took when I was an undergraduate student in [country]. It was difficult because I was terrible at English. Before I went to the(D) university, I learned only reading and writing in English. The topics of the materials I had learned were extremely difficult and boring. It was because I had learned English just for the [country's] SAT(P) which is the test that high school third year(SPG) third-year(WO) students take to go to university. I was a student who studied (native language) secretly in English class, since I hated English then. I took the English 1 class right before I entered my university. I lost(VF) the goal to keep learning English(Awk I'm not sure what you mean here) because I studied English only for the [country's] SAT and I wanted to avoid the class, but it was mandatory(SS) It was hard for me to get used to learn(VF) English in (*D) environment that everyone speaks(VT spoke) and listens(VF) in English. The worst part of the class was a short presentation. Even though I had a chance to pick (*D) topic which I liked, I kept postponing preparing it. Finally, on the presentation day, I felt so nervous and wanted to run away.</p> | |
| Rewrite 2 | <p>The most difficult course I have ever taken was English 1, which I took when I was an undergraduate student in [country]. It was difficult because I was terrible at English. Before I went to university, I learned only reading and writing in English. The topics of the materials I had learned were extremely difficult and boring. It was because I had learned English just for the [country's] SAT, which is the test that third-year high school students take to go to university. I was a student who studied (native language) secretly in English class, since I hated English then(WC at that time). I took the English 1 class right after I entered my university. I had no motivation to keep learning English because I studied English only for the [country's] SAT when I was a high school student. So(WC "so is not an academic way to begin a sentence) I wanted to avoid the class, but it was a(D x) one of the mandatory classes which I must(VT needed) (*PP to) enroll (*PP in) to graduate (*PP from) the(D x) university. It was hard for me to get used to learning English in the(D an) environment that everyone spoke and listened in English. The worst part of the class was a short presentation. Even though I had a chance to pick the(D a) topic which I liked, I kept postponing preparing it. Finally, on the presentation day, I felt so nervous and wanted to run away.</p> | done |
| Prompt 2: | Name a controversial topic in your field and delineate your position. | |
| | <p>One of (D) controversial topics in teaching and learning second language(S/Pl) is whether writing feedback is helpful for students or not. There has(SV) been many studies to figure this out(P) and some scholars argued(VT) that giving feedback to(PP on) students' writing would not only reduce but also benefit for students' (CS). They studied students' into how receive and make feedback only to one</p> | |
| <p>+ Paragraphs Symbols Error Tally Log Edit Log</p> | | |

Appendix C: Error Codes

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