TESL-EJ Teaching English as a Second or Foreign Language
The Electronic Journal for English as a Second Language

Can Reading Aloud Replace Teacher Feedback? A Comparative Analysis of Proofreading in ESL Writing

February 2025– Volume 28, Number 4
https://doi.org/10.55593/ej.28112a5

Grant Eckstein
Brigham Young University, USA
<grant_eckstein@byu.edu>

Ying Suet Michelle Lung
The University of Utah, USA
<ysmichelle.lung@gmail.com>

Natasha Gillette
Brigham Young University-Hawaii, USA
<Natasha.Gillette@byuh.edu>

Abstract

Students are often encouraged to proofread their writing by reading it aloud. Presumably, this will allow writers to correct local errors. Yet even though this strategy may be effective for native speakers, there is little empirical evidence of its benefit among second language writers. Therefore, we wondered how many errors second language students could correct through reading aloud and how that compared when receiving teacher feedback. In this study, 60 ESL students composed four in-class, 10-minute essays over two weeks. Half of the students revised their essays after receiving teacher feedback; the other half revised after reading aloud. All errors on initial and revised essays were tallied and normalized. Results showed that reading aloud affected some surface errors but that teacher feedback significantly outperformed reading aloud. We recommend teachers use reading aloud as a supplementary or preliminary strategy for correcting surface errors but not as a replacement for expert feedback provision.

Keywords: proofread; reading aloud; out loud; WCF; L2; error correction; feedback

A popular proofreading strategy that writing center tutors and composition teachers encourage students to use is reading aloud (RA). In a study of researcher writing strategies, Bakla and Karakaş (2022) showed that 13 out of 31 academic writers mentioned using RA as a writing or revision strategy; moreover, in a follow-up survey, it was the third most frequently reported strategy among writers. The RA process typically involves students reading their writing aloud either to themselves in private or to a peer, tutor, or instructor to eliminate errors or improve

writing quality. Elbow (2012) focused an entire book on RA and explained as part of his larger argument that revising aloud addresses many grammar mistakes originating from carelessness or problems of meaning. Elbow further explained that the combination of RA and revising aloud assists writers in noticing errors as writers become attuned to "*hear* a lapse in logic" (p. 225). Hartwell (1985) stated that when students revise this way, they will "correct in essence all errors of spelling, grammar, and, by intonation, punctuation" (p. 121). And writers seem to agree. Rowe (2010) interviewed and observed experienced writers who used RA as a regular part of their revision process. Many attributed their writing success to this articulatory and cognitive activity. They regarded RA as a prominent and vital part of revision in writing because it increased their awareness of global-level and sentence-level revisions.

Because of the popularity of RA generally, it has been adopted and encouraged among English as a Second Language (ESL) writers too. Gibson (2008), for instance, found that 18 out of 22 ESL writers used RA to review their work. Yet there is some question of whether ESL writers benefit from this practice, with limited and contradictory evidence to that effect (Çetinkaya, 2020; Hedgcock & Lefkowitz, 1992; Tseng, 2014). Hedgcock and Lefkowitz (1992) found that second language (L2) writers of French improved in content, organization, and vocabulary as a result of RA, while those who received only teacher-written feedback worsened in these areas but improved in grammar. On the other hand, Tseng's (2014) research suggests that ESL writers struggle to identify global issues during RA but that as their proficiency in English increases, they become more adept at correcting local errors. Similarly, Çetinkaya (2020) found that ESL writers were more successful at identifying surface-level errors than semantic errors in RA revision.

While RA seems promising, research makes it unclear whether RA is as good as teacher feedback (TF) at helping students catch and correct surface errors. With the increasing popularity of independent studies and online courses, not all students have the luxury of receiving feedback from teachers and thus may increasingly turn to RA as a replacement for TF. We recognize that some experts may argue that RA and TF are not comparable since writing teachers have significantly more language knowledge to bring to bear in identifying errors and since teachers are more likely than their students to approach student writing from a fresh perspective; however, it is still important to know whether RA can be as useful in correcting some or possibly all surface errors as some researchers have predicted (e.g., Hartwell, 1985) and whether RA is an economical and convenient replacement for TF so that teachers can eliminate some of their feedback burden and help students thrive as independent revisers. Therefore, we designed a writing study in a typical writing class setting to compare surface revision when students used RA versus when they used TF.

Theoretical Framework

Despite the ubiquitous application of RA for self-revision in writing centers and elsewhere, there appears to be no theoretical basis to explain its operation (S. Polio, personal communication, 14 June 2023). We propose that when reading aloud is used as an editing strategy, it is largely a form of proofreading as opposed to composing or writing (Azeez, 2020; Davis, 1995). The basic purpose of proofreading, and as such, its definition, is to identify and correct "typographical, linguistic, coding or positional errors or omissions" (Chartered Institute of Editing and Proofreading, 2020). The processes underlying proofreading differ little from reading; instead of processing letters and words automatically, with comprehension as the immediate goal, a proofreader intentionally resists automatic word recognition and semantic

meaning-making processes typical in fluent reading in order to prioritize orthographic processing of letter strings and thus catch errors. As Pilotti et al. (2012) explained:

> Proofreading of word errors requires that letter processing be completed for each letter string and that the outcome of this processing be compared with any lexical interpretation generated in response to the letter string. (p. 643)

By this understanding, proofreading differs from reading because it is slower, more effortful, and focused on orthographic and syntactic detail that is hardly perceptible when reading quickly (Larigauderie et al., 2020; Schotter et al., 2014). Such an explanation of proofreading does not distinguish silent from oral proofreading, which may be unproblematic since oral and silent reading rely on similar processes but differ in the extra cognitive demands associated with oral pronunciation of words (Hale et al., 2007). It is this extra cognitive effort that may make reading aloud an attractive strategy for proofreading since it inherently facilitates a slower reading rate (Cushing & Bodner, 2022).

We argue, along with Koda (2007), that all reading (and thus all proofreading) "builds on oral language competence" (p. 1). This is because auditory knowledge is foundational for literacy development (Willis, 2008). That is, both silent and oral proofreading utilize phonological processes wherein readers "hear" the words in their mind. To account for the oral component of reading, we propose a phonological processing model of proofreading. This hypothesized model draws theoretically from a dual-route approach to reading (Coltheart, 2005), which presumes that when encountering printed words, readers typically convert the printed form to its phonological form, meaning, readers "hear the words" either aloud or in their mind (Coltheart, 2005). This conversion process can take place in one of two routes: by decoding words at the letter level through sound-letter correspondences or by whole word phonological recognition of familiar words (Zorzi et al., 1998). The latter process, it is presumed, is intentionally restricted in proofreading (Pilotti et al., 2012). In proofreading, as with reading, readers access cognitive systems such as the mental lexicon (Coltheart et al., 2001) and a syntactic system (MacKay et al., 2021; Schotter et al., 2014), which pull existing phonological, morphological, and syntactic information from long-term memory.

In a typical reading task, the output of the dual route approach to reading is comprehension. But in our proofreading model, the output is phonological information which is then subjected to noticing processes for the purposes of editing (see Davis, 1995). As illustrated in Figure 1, proofreaders transfer orthographic forms to oral phonological output by means of letter decoding. The suppression of whole word processing while preparing to articulate words aloud serves as a resistor on the system. As the text is transferred to phonological output, the reader activates cognitive systems associated with lexical and syntactic comprehension. Readers may notice discrepancies between the written form in the text and their own phonological production, a proofreading phenomenon observed early on by Bartholomae (1980) and later referred to as a *production effect* by MacLeod et al. (2010). This may occur thanks to misspellings, aberrant syntactic forms, unusual word choice, missing or misplaced punctuation, or other mismatches between the written form and the phonological form. But readers will not notice every error. The effectiveness of noticing is mitigated by the reader's cognitive systems (Schotter et al., 2014); textual features (e.g., errors in more frequent, shorter, and predictable words are harder to spot [Pilotti et al., 2012]); and noticing capacity, which can be constrained by such things as reader age, reading/proofreading experience, difficulty or familiarity of the text, language interference, error type, and so on (Shafto, 2015). Where noticing does occur,

and the reader is concerned with obvious discrepancies, an error correction process ensues which either adjusts the phonological form to align with the written form or adjusts the written form to align with the phonological form. If the second path is taken, this leads to textual emendations, or editing.

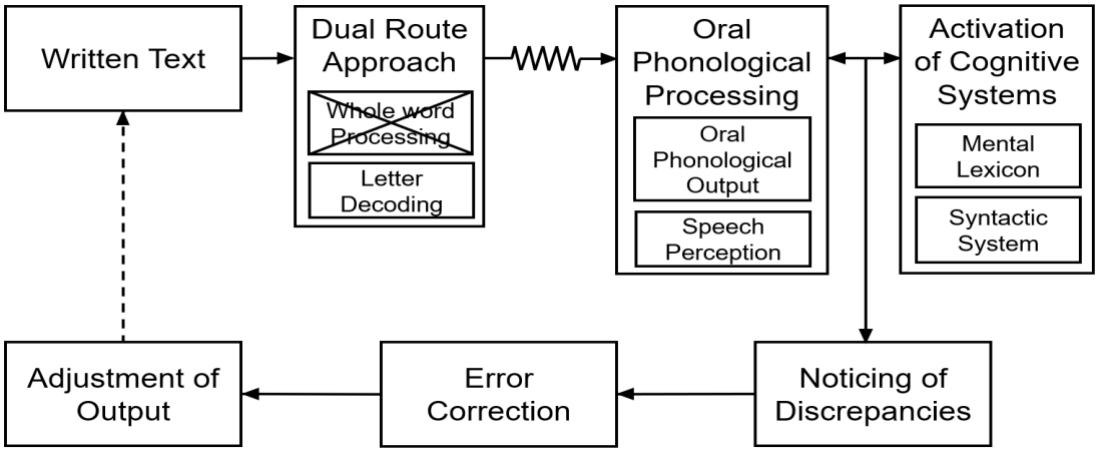

Figure 1. Phonological Processing Model of Proofreading

In contrast to this model, the instructor-initiated written corrective feedback model is, in our view, a faster, simpler process. In this model, readers edit their writing by potentially engaging with the text, though this step is not strictly necessary as readers may attend immediately to error correction marks or in-text corrections. As with the phonological processing model, readers also activate cognitive systems that may trigger noticing strategies or may simply direct readers to error correction.

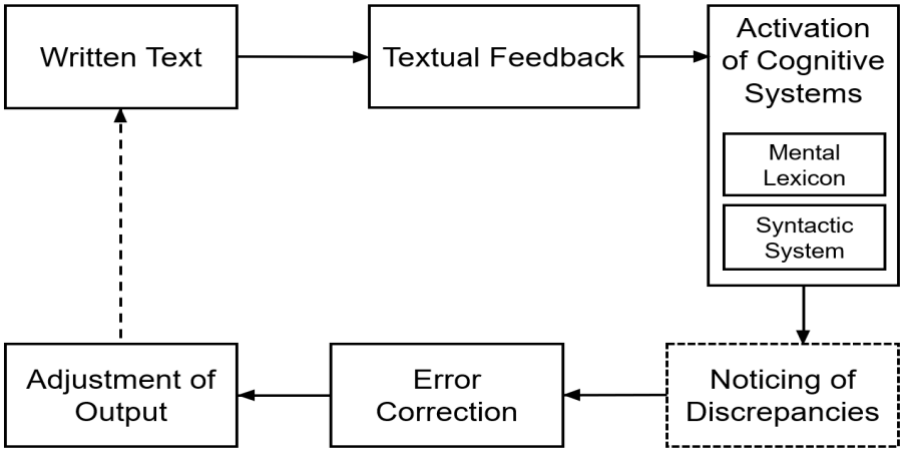

Figure 2. Written Corrective Feedback Model

We hypothesize that both models lead to error correction. We predict that the slower, more resource-intensive phonological processing model of proofreading leads to greater long-term gains in self-editing development because of the forced interaction with each word. However, if the written corrective feedback model were initiated by a language expert, we postulate that it could produce superior results in both the short and long term because the editor would have access to linguistic knowledge beyond that of the reader. We hypothesize that only when a self-editor with expert linguistic knowledge or native-like intuition reads aloud do both the

phonological processing model of proofreading and the written corrective feedback model produce similar short- and long-term results.

Literature Review

RA is one self-editing approach that writing teachers and especially writing center administrators have encouraged for decades (see Block, 2016; Elbow, 2012; Perl, 2006; Ryan & Zimmerelli, 2006). In a writing center context, Powers (1993a) referred to RA as "that seemingly undebatable practice of asking writers to read their own drafts aloud" (p. 3). She went on to explain its value: "to encourage writers to self-edit, to assess voice, to assume ownership, to hear punctuation" (p. 3). Even earlier, Hartwell (1985) explained that when students read aloud, they verbally correct errors in their text "without noticing that what they read departs from what they wrote" (p. 121). Such experiential or intuition-based advice continues to be doled out in recent guidebooks and internet tips for proofreading (Cushing & Bodner, 2022; Hampton, 2019). Presumably, reading aloud produces an encoding effect, which makes words more salient and may facilitate the noticing of errors (MacLeod & Bodner, 2017; Robinson et al., 2019).

To illustrate how this works in the real world, Ferris (2011) related a family anecdote about reading aloud in which her daughter was struggling with verb tenses in Spanish as a foreign language. Ferris asked her daughter to read a recent essay draft aloud while focusing on verbs. Ultimately, only two sentences had incorrect verb forms, one which could be corrected by checking in a dictionary and another which was identified when Ferris later read the sentence aloud to her daughter. Both errors could be managed by returning to a grammar textbook or dictionary to review specific rules. Ferris's work provides one salient example of how RA could work and why writing teachers recommend that students read their own writing aloud, even in a foreign language.

While some researchers and teachers view RA positively in an ESL context (Çetinkaya, 2020; Hedgcock & Lefkowitz, 1992; Tseng, 2014), other researchers have approached RA more cautiously. Matsuda and Cox (2009) acknowledged that "because ESL writers often have not internalized some of the rules of grammar, they are often not able to identify errors on their own by, for example, reading the text aloud" (p. 44), something they agree is typically useful for native English (L1) writers and commonly practiced in writing centers. For their part, writing center tutors and administrators have recognized similar limitations but, unwilling to abandon an oral component, have instead suggested that a tutor read the student's writing aloud for him or her (Purcell, 1998), just as composition teachers might ask students use text-to-speech technology (Malin, 2019) or to read a peer's text in order to avoid what Borrowman (2004) called "the trauma of reading aloud," where students frequently feel shame, apologize profusely, and even resist reading their own drafts aloud as part of the revising process.

Despite its prevalence as a recommended revision technique, RA seems to be heavily based on beliefs and intuitions rather than research. Many writing center tutors encourage students to read their papers aloud because this practice has been passed on for years (Lott, 2016; Rafoth, 2017; The Writing Center, University of North Carolina at Chapel Hill, 2024). Moreover, RA is promoted as if it had been critically validated, yet this does not appear to be the case. Only a few studies have empirically evaluated RA in the L1 writing field (Cushing & Bodner, 2022; Riefer, 1993; St. John, 2004). The earliest study we found, by Riefer (1993), showed that RA was effective in locating spelling errors, though it was not more effective than reading silently.

In another study, St. John (2004) had undergraduates complete an initial essay draft, then use RA revision in making changes. Results showed that students largely made surface-level revisions. Cushing and Bodner (2022) more recently demonstrated that RA led to more error detection than reading in a disfluent font or reading silently. These three studies, while positive toward RA, hardly provide the empirical justification one would expect for an intervention so universally encouraged. Furthermore, it is questionable whether RA—an approach ostensibly developed for L1 writers to capitalize on their native intuition of English—is applicable to L2 writing.

For its part, RA has been recommended for L2 writers in writing center contexts, high school English writing classes, and first-year college writing programs. Gibson (2008) investigated the benefits associated with the use of RA for language learning and identified usages like diagnostic purposes, pronunciation and prosody evaluation, and writing revisions. In addition, she also interviewed teachers (12 L1 teachers and 15 ESL teachers) and English language learners (seven students from various language backgrounds) to explore how and why RA was being used in language learning. The results showed that 82% of the ESL participants used the RA strategy in their private study, indicating that most of the students who took charge of their learning used this strategy to evaluate their work. Additionally, the participants saw RA as a critical pronunciation activity and a way to enhance their engagement in learning.

In terms of correction approaches, Hedgcock and Lefkowitz (1992) investigated the effects of RA among 30 students producing multiple-draft essays in their second language, which was French. Half the group received only written feedback from the teacher on intermediate drafts, while the other half read their drafts aloud to two partners. The RA group significantly outperformed the written feedback group on content, organization, and vocabulary; meanwhile, the written feedback group worsened in these categories but improved considerably in grammar. The results suggest that RA can lead to rhetorical improvements in L2 writing.

Tseng (2014) studied 28 college-level ESL writers in Taiwan who were divided into basic, intermediate, and advanced English proficiency groups and found conflicting results. All participants wrote two multi-draft essays and subsequently read their first drafts aloud twice to an instructor; both students and teacher marked local and global errors. Results showed that writers only identified a few global errors for revision regardless of proficiency level and, in fact, only identified about a quarter of the global errors their teachers identified. In contrast, students could generally detect many local problems after reading their papers out loud. Results showed that proficiency matters in detecting local problems; intermediate and advanced writers could respectively identify 51% and 89% of local problems while basic level writers could only identify 20%. The results suggest that ESL writers struggle to identify global issues while reading aloud, but as the writers' proficiency increases, so does facility in correcting local errors.

Similarly, Çetinkaya (2020) conducted research with 50 fourth-year university students in Turkey. All participants were tasked with two single-draft essays where they revised the first one silently and the second one aloud. They also had to mark the location of any error they saw during the process. Results showed that participants succeeded in detecting more surface and semantic errors in read-aloud revision than in silent revision. Nevertheless, RA was more functional in detecting surface-level errors, while silent revision was more useful in the semantic dimensions.

The practice of RA has long been advocated by writing teachers and writing center administrators due to its perceived benefits in self-editing and error detection. While some scholars support its efficacy in both L1 and L2 writing contexts (Block, 2016; Cushing & Bodner, 2022; Elbow, 2012; Ferris, 2011; Hartwell, 1985; MacLeod & Bodner, 2017; Perl, 2006; Robinson et al., 2019; Ryan & Zimmerelli, 2006), others question its practicality and usefulness among second language learners, citing the lack of intuitive grasp and dependence on native-speaker competence (Borrowman, 2004; Çetinkaya, 2020; Hedgcock & Lefkowitz, 1992; Malin, 2019; Matsuda & Cox, 2009; Powers, 1993a; Purcell, 1998; Tseng, 2014). For instance, Cogie et al. (1999) explained forcefully that "the read-aloud method for discovering sentence-level errors, frequently productive for native speakers, provides little help to ESL students who lack the ear to hear their own error" (p 7). And Powers (1993b) observed in her own writing center that "neither reading aloud nor editing by ear appears to work for the majority of ESL writers we see" (p.42). This raises questions about the productivity of RA for L2 writers in terms of error correction. Therefore, our study was motivated by the following research questions:

1. To what extent does RA lead to surface-level proofreading improvement compared to TF for ESL writers?
2. To what extent does English language ability level affect these proofreading results?

Method

Participants

This study included 60 adult ESL learners (treatment: $n = 30$, control: $n = 30$) from four intact classes, two in each condition. Student information is included in Table 1. Classes were taught by three different teachers. They met four days per week, 65 minutes per class period over 13 weeks.

All participants were enrolled in an Intensive English Program (IEP) associated with a large teaching university in the western US. Students at the IEP are assigned to specific tracks based on the American Council on the Teaching of Foreign Languages (ACTFL) proficiency scale (ACTFL, 2012). Students participating in this study came from two tracks whose ACTFL proficiency scales were intermediate-mid to intermediate-high and advanced-low to advanced-mid, equivalent to levels B1 and B2 of the Common European Framework of Reference (CEFR) (ACTFL, 2016). One class from each track was randomly selected to be the treatment condition.

Table 1. Control and Treatment Groups Participant Information

	Experimental Group (Intermediate: $n = 16$, Advanced: $n = 14$)		**Control Group** (Intermediate: $n = 16$, Advanced: $n = 14$)	
Gender	15 female, 15 male		15 female, 15 male	
Age	18–35 (Mean: 23)		18–53 (Mean: 26)	
Native language	3 Japanese 2 Korean 1 Papuan	2 Portuguese 22 Spanish	1 Chinese 3 Japanese 2 Korean	2 Portuguese 21 Spanish 1 Tongan

Procedures and Instruments

Following ethics board approval, the study was conducted over the course of two weeks, during which participants wrote four sets of 10-minute paragraphs (Draft 1 and Draft 2 of each). The writing process for each set of writing tasks for each condition was the same. Both treatment and control groups were tasked with responding to an in-class, 10-minute writing prompt on intentionally general topics such as "too much freedom" and "the value of homework"; the generality of the topics allowed students to be flexible in developing their thoughts. All participants typed their responses on a quiz utility created on Canvas, an online learning management system; spelling and grammar check was disabled. After submitting their original response (Draft 1), the treatment students were asked to immediately read their writing aloud to themselves, revise the writing for five minutes, and submit their revised response (Draft 2) in class to a second quiz utility. Meanwhile, in the control group, students did not read their Draft 1 paragraphs out loud but instead received teacher feedback in the form of codes designed for local error correction (see Hartshorn et al., 2010 for details of the coding scheme) on the next day of instruction, at which point the control group spent five minutes in class correcting their errors and then submitted their revised writing (Draft 2) on Canvas. Figure 3 shows the research procedures for the first of two weeks; the same procedure was repeated during the second week.

One reviewer of a previous version of the article pointed out that the timing of feedback may have an effect on results, and we concur that the difference in feedback timing for the two groups was slightly asymmetrical and could have been better controlled by, say, delaying RA one day. As Daneman and Stainton (1993) reported, immediate proofreading (20 minutes after composing) was less effective than proofreading delayed by two weeks. However, Pilotti et al. (2006) found no difference in proofreading speed or accuracy between an immediate (10-minute delay) and delayed (40-minute delay) proofreading task of familiar text. Since immediacy of feedback used in the classes of our study was a defining feature of the intervention (see Eckstein et al., 2020; Hartshorn et al., 2010) and because we did not want to wait two weeks to instigate feedback, we chose to use the present design.

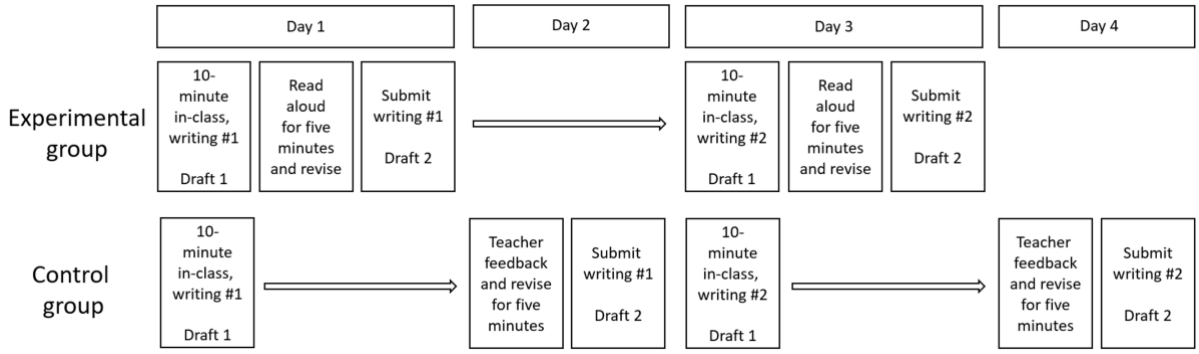

Figure 3 Experiment Procedures for Both Conditions

Data Coding and Analysis

Three researchers participated in data coding; two held PhDs in linguistics or education, and all three held master's degrees in TESOL. One of the researchers had extensive experience using the error coding scheme and introduced it to the other coders (see Table 2 for coding details; Eckstein & Ferris, 2018). After one hour of training on nine error measures using

Dedoose, a software package developed for qualitative data coding, the researchers individually practiced coding on a set of 10 paragraphs. Following norming and further training, the researchers reached an interrater reliability Kappa score of over 85% on a secondary subsample of 20 essays. The full set of essays was then divided among the three researchers such that every essay was coded by a primary and secondary coder so that all essays were double-coded. When researchers disagreed on a code application, they would discuss and come to an agreement about the language structure or code and thus resolve all disagreements in this fashion.

The total number of errors and errors in each category for each draft were tallied, resulting in nine scores (e.g., mechanics, pronoun usage, punctuation, etc.) for all first and second drafts for all four 10-minute paragraphs per participant. Scores were then normalized to ensure comparability across paragraphs and participants. This was done by dividing the error count by the number of words in the paragraph and then multiplying by a constant of 135, which was the average word count of all paragraphs in the study.

A repeated measures ANOVA was employed to analyze language errors across first and second drafts of timed paragraph writing, comparing two student groups—those revising with teacher feedback and those revising by reading aloud—across intermediate and advanced English proficiency levels. The analysis featured one within-subject factor, draft (draft 1 vs. draft 2), and two between-subject factors: group (teacher feedback [TF] vs. reading aloud [RA]) and class (intermediate vs. advanced). Thus, the repeated measures ANOVA included draft as the repeated measure and group and class as between-subject factors, resulting in a single 2x2x2 model incorporating draft × group × class interactions. With two levels per factor, simple main effects (e.g., draft 1 vs. draft 2) were evident from the RM ANOVA. Significant interaction effects (e.g., draft × group) were further examined through post hoc comparisons (Larson-Hall, 2010). The data met the assumptions of RM ANOVA including sphericity and normality. We set the alpha level at .005 (.05/10) to account for potential Type I errors, given we analyzed ten measures for each paragraph.

Table 2. Error Categories Used for Coding

Major error category	Brief description
Punctuation	Missing/unnecessary commas, semicolons, apostrophes, quotation marks
Mechanics	Spelling/typing errors; capitalization errors; missing/incorrect hyphens on compounds
Nouns/noun phrases	Missing/unnecessary/incorrect plural markers, possessive markers, articles/determiners
Subject-verb agreement	Error in the noun/verb phrase
Verbs/verb phrases	Incorrect tense/aspect; passive voice incorrectly formed; modal auxiliary incorrect
Sentence structure	Run-ons; comma splices; fragments; word order; missing/unnecessary words
Word form	Wrong word form for context, including verb form errors not covered in verb category
Pronoun usage	Unclear pronoun reference/incorrect pronoun form
Incorrect word choice	Any lexical error, including preposition errors

Results

The purpose of this study was to explore how reading aloud in the writing revision process relates to proofreading performance . We also wanted to compare results across students' language ability level. For a full presentation of means and standard deviations for all groups and measures, see Appendix A; statistical output for the full RM ANOVA model is listed in Appendix B. Our analysis revealed a main within-subjects effect of draft with p-values less than .001 for all variables (save pronoun errors for which there were too few observations).

This indicates that revised drafts had fewer errors than first drafts overall. Nevertheless, effect sizes in terms of eta-squared ranged from .003 to .023, indicating only negligible to small effects since typically, eta-squared values are considered negligible < .01 and small where .01 ≤ η2 < .06 (Norouzian & Plonsky, 2018). Table 3 shows all significant values for draft while Figure 4 illustrates the difference in total errors between drafts graphically.

Table 3. Main Within-Subjects Effects for Draft

	Draft	*n*	*M*	*SD*	*f*	*df*	*p*	η^2
Mechanics	1	232	4.17	4.1	139.2	210	< .001	0.022
	2	214	2.36	3.3				
Nouns/noun phrases	1	232	2.90	3.0	43.23	210	< .001	0.007
	2	214	2.18	2.6				
Pronoun usage	1	232	0.19	0.5	3.29	210	0.071	0.001
	2	214	0.17	0.5				
Punctuation	1	232	1.90	1.8	17.69	210	< .001	0.003
	2	214	1.55	1.6				
Sentence structure	1	232	4.77	3.1	67.38	210	< .001	0.007
	2	214	4.04	2.9				
Subject-verb agreement	1	232	0.70	1.1	20.02	210	< .001	0.003
	2	214	0.50	1.0				
Verbs/verb phrases	1	232	1.14	1.8	19.69	210	< .001	0.002
	2	214	0.86	1.4				
Word form	1	232	1.24	1.4	22.57	210	< .001	0.006
	2	214	0.92	1.3				
Word choice	1	232	2.36	2.2	37.55	210	< .001	0.004
	2	214	1.95	2.1				
Totals	1	232	19.38	10.7	306.4	210	< .001	0.023
	2	214	14.53	10.1				

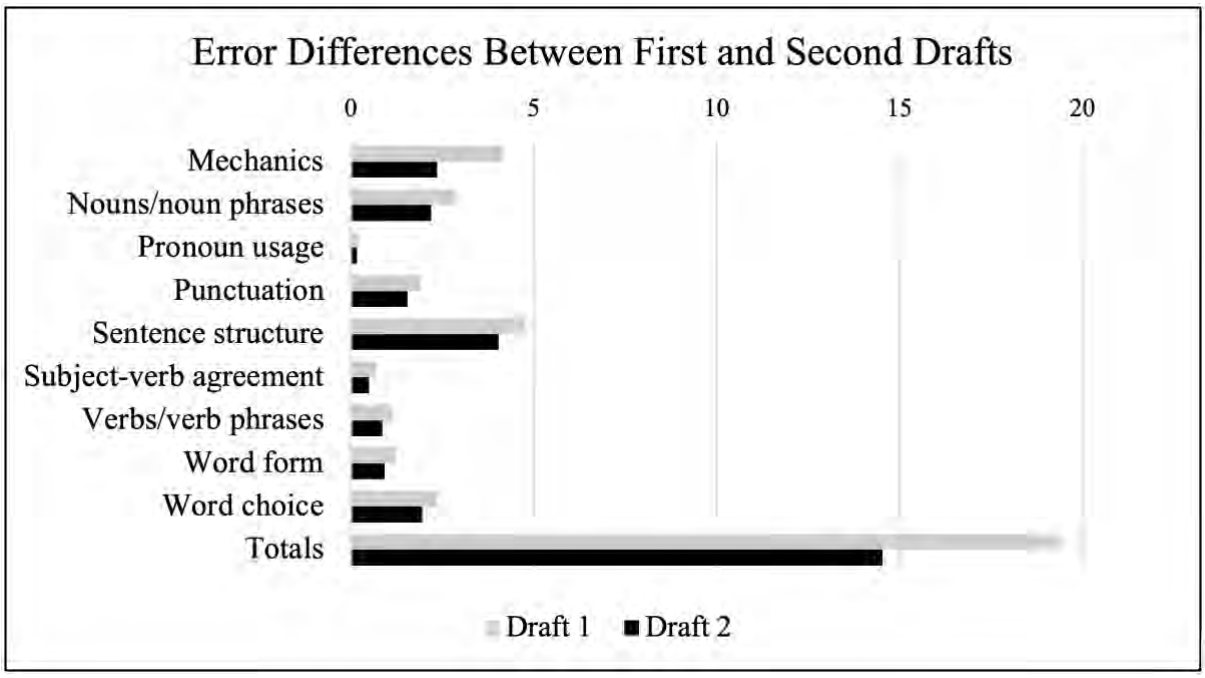

Figure 4. Errors across all papers for Draft 1 and Draft 2

Similarly, we found a between-subjects main effect for class level which was significant for all variables except pronouns as shown numerically in Table 4 and graphically in Figure 5. This indicates that advanced learners made fewer errors than intermediate learners, a finding that should not be surprising to those familiar with second language development. Eta-squared effect sizes ranged from .011 to .058 except for total errors, which was .101, suggesting that class level accounted for about 10% of the total error variance, a medium effect size.

Table 4. Main Between-Subjects Effects for Class

	Class	*n*	*M*	*SD*	*f*	*df*	*p*	η^2
Mechanics	AA	244	4.37	4.4	25.11	210	< .001	0.036
	UP	202	2.02	2.5				
Nouns/noun phrases	AA	244	3.44	3.1	32.84	210	< .001	0.047
	UP	202	1.49	2.0				
Pronoun usage	AA	244	0.23	0.6	3.48	210	0.063	0.005
	UP	202	0.12	0.4				
Punctuation	AA	244	2.00	1.9	7.2	210	0.008	0.011
	UP	202	1.41	1.3				
Sentence structure	AA	244	5.23	3.4	23.7	210	< .001	0.035
	UP	202	3.45	2.2				
Subject-verb agreement	AA	244	0.83	1.2	12.39	210	< .001	0.019
	UP	202	0.33	0.6				
Verbs/verb phrases	AA	244	1.42	2.0	21.5	210	< .001	0.031
	UP	202	0.50	0.8				
Word form	AA	244	1.36	1.5	13.5	210	< .001	0.02
	UP	202	0.76	1.1				
Word choice	AA	244	2.94	2.4	40.36	210	< .001	0.058
	UP	202	1.22	1.4				
Totals	AA	244	21.81	11.0	74.3	210	< .001	0.101
	UP	202	11.29	6.8				

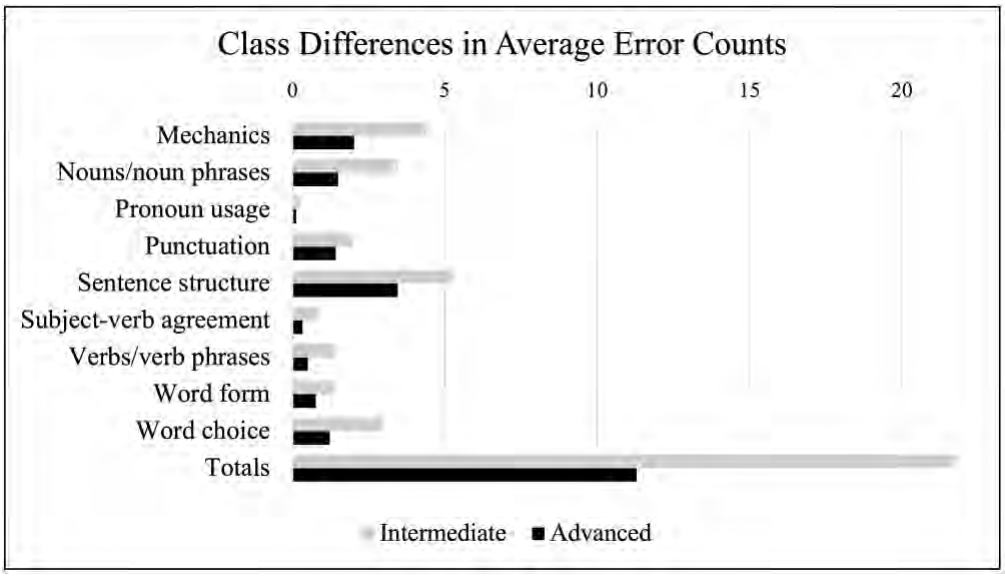

Figure 5. Errors across all paragraphs for intermediate and advanced learners

We did not find significant between-subject main effects for group, except for verbs/verb phrases ($F(210) = 15.70$, $p < .001$, $\eta^2 = 0.023$) meaning that students assigned to the teacher

feedback and reading aloud groups did not differ overall except in verb errors in which students in the teacher feedback group had more verb errors. At this level, the result indicate that RA and TF groups were well balanced. Confirming these main effects, we proceeded with our analysis of interactions to answer our specific research questions.

Effect of Error Correction Approach on Proofreading Performance

Our first research question examined the extent to which reading aloud led to greater proofreading performance compared to teacher feedback from first to second drafts. When examining the interaction between draft and error correction group, several variables demonstrated that teacher feedback was superior to reading aloud. Table 5 shows draft x group interactions with associated statistics. For instance, in the variable of mechanics, there were about four errors per 10-minute paragraph in the first draft. After teacher feedback, student mechanical errors reduced to 1.81, whereas students with reading aloud feedback reduced their errors to just 2.86. This represents an average reduction of nearly 2.5 mechanical errors per paragraph for the teacher feedback group, compared to about 1.2 errors for the reading aloud group. A post hoc analysis showed that both reading aloud and teacher feedback led to significant reductions in mechanical errors. In addition, nouns, sentence structure, subject-verb agreement, word choice, and total errors showed significant interaction effects.

Table 5. Interaction effects of draft and error correction approach on proofreading performance

		Teacher Feedback			Reading Aloud						
	Draft	**N**	**M**	**SD**	**N**	**M**	**SD**	*f*	*df*	*p*	η^2
Mechanics	1	116	4.28	4.43	116	4.07	3.83	18.90	210	<.001	0.003
	2	102	1.81	3.32	112	2.86	3.26				
Nouns/noun phrases	1	116	3.00	3.16	116	2.81	2.84	12.90	210	<.001	0.002
	2	102	1.91	2.55	112	2.43	2.63				
Pronoun usage	1	116	0.14	0.50	116	0.25	0.52	0.31	210	0.581	0.000
	2	102	0.12	0.44	112	0.21	0.50				
Punctuation	1	116	2.11	1.78	116	1.69	1.70	1.30	210	0.256	0.000
	2	102	1.63	1.42	112	1.48	1.67				
Sentence structure	1	116	5.26	3.20	116	4.28	3.01	24.83	210	<.001	0.003
	2	102	4.18	3.01	112	3.91	2.78				
Subject-verb agreement	1	116	0.78	1.20	116	0.62	0.98	16.78	210	<.001	0.003
	2	102	0.41	0.94	112	0.58	0.98				
Verbs/verb phrases	1	116	0.72	1.46	116	1.55	1.97	0.71	210	0.400	0.000
	2	102	0.43	1.06	112	1.26	1.54				
Word form	1	116	1.29	1.62	116	1.19	1.15	4.57	210	0.034	0.001
	2	102	0.82	1.40	112	1.01	1.14				
Word choice	1	116	2.49	2.30	116	2.23	2.18	11.97	210	<.001	0.001
	2	102	1.87	2.16	112	2.02	2.00				
Totals	1	116	20.05	11.04	116	18.70	10.34	65.08	210	<.001	0.005
	2	102	13.19	10.46	112	15.75	9.60				

We therefore conducted post hoc analyses on significant interactions, which showed significant reductions of errors over time just for those receiving teacher feedback. In six cases in

particular, teacher feedback resulted in a significant decrease in errors from draft 1 to draft 2 while reading aloud did not. These variables included nouns $p<.001$, $d=.378$, punctuation $p=.002$, $d=.299$, sentence structure $p<.001$, $d=.348$, subject-verb agreement $p<.001$, $d=.341$, word form $p<.001$, $d=.307$, and word choice $p<.001$, $d=.279$. These six can be seen charted in Figure 6.

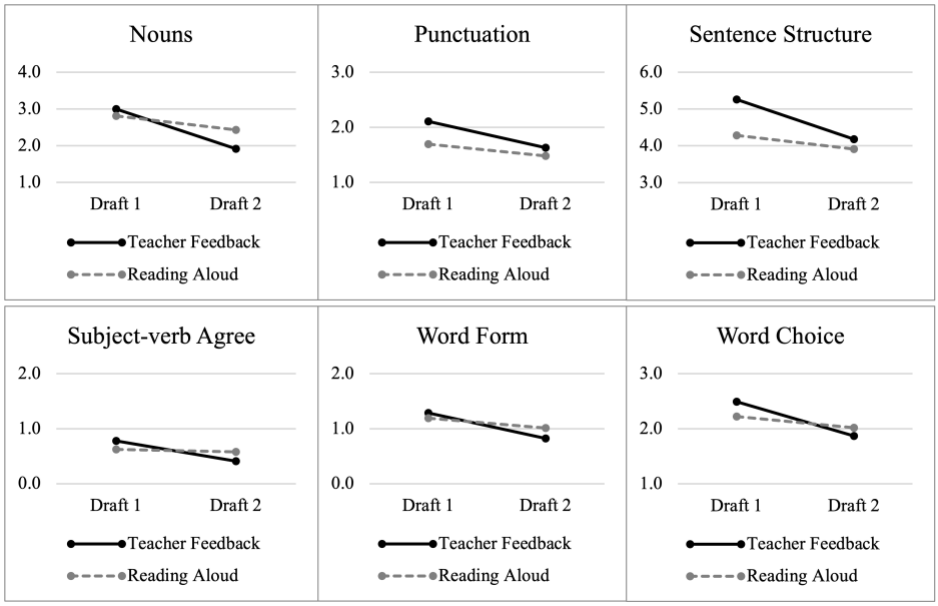

Figure 6. Error reduction where TF is significant while RA is not

Total errors showed a pattern similar to that of mechanics in that both reading aloud and teacher feedback resulted in significant reductions in errors, but the difference was striking. When reading aloud, students were only able to correct about three errors per paragraph $p<.001$, $d=.295$, but with the help of teacher feedback, students corrected nearly seven errors per paragraph $p<.001$, $d=.639$. Clearly, reading aloud was beneficial for reducing total errors over time, but receiving teacher feedback was far superior. See Figure 7.

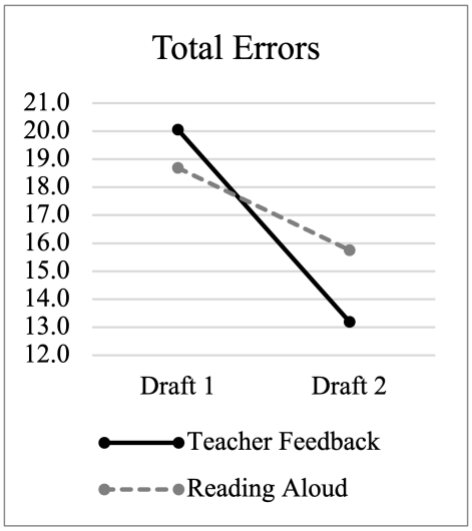

Figure 7. Total error reduction where TF is more beneficial than RA

Effect of Proficiency Group and Error Correction Approach on Proofreading Performance

For the second research question, we examined whether proofreading performance (draft) differed between intermediate and advanced learners (class) for the different feedback approaches (group). Although advanced learners outperformed intermediate learners with fewer errors in all categories (as shown in main between-subjects effects above), when accounting for error correction approach and language ability level across drafts, only the error category of punctuation showed significant interaction effects at the adjusted alpha level of .005 ($F(1, 1) = 9.69$, $p = 0.002$, $\eta^2 = .002$) as seen in Figure 8. A post-hoc analysis revealed that only teacher feedback for the intermediate class was significant from first draft (m = 2.38, sd = 2.03) to second draft (m = 1.60, sd = 1.44) ($p < .001$). All other punctuation p-values were > 0.33. In other words, we found no evidence that students in different language ability levels benefited more or less from RA or from TF. The one exception was punctuation in which intermediate-level students benefited more from TF than advanced students.

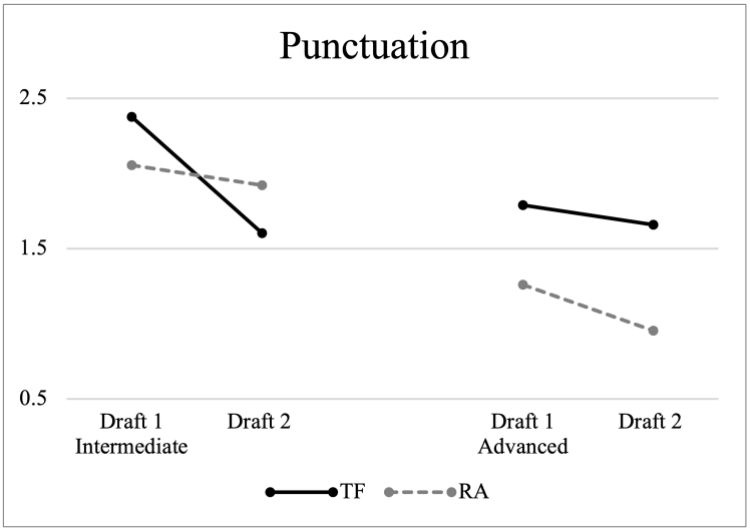

Figure 8. Punctuation errors shown for draft x group x class

Discussion and Conclusions

The results of the study indicate that any kind of feedback appears to support proofreading. The RA method was effective in eliminating an average of three errors per 10-minute paragraph, which corroborates claims made by Hartwell (1985) that RA can effectively correct errors of spelling, grammar, and punctuation by intonation. On the other hand, TF eliminated an average of seven errors per 10-minute paragraph, and it was significantly better than RA in addressing six of the nine error categories, specifically nouns, punctuation, sentence structure, subject-verb agreement, word form, and word choice. In these cases, students were able to catch mistakes more effectively after receiving TF than RA. This confirms what Gibson (2008) said about RA's limited and contradictory evidence to support its use among L2 writers. Based on this research, we tend to agree with the observation of some writing center practitioners (i.e., Cogie et al., 1999; Powers, 1993a,b) that RA is only minimally effective for L2 writers, likely because the process of "hearing" errors requires native-like intuition, an extreme grasp of specific grammatical principles, or teacher-supported RA practices. While L1 writers generally

have an intuition of the linguistics rules, L2 writers need explicit instruction or training to identify and correct errors while reading their writing out loud.

Advanced learners were shown to be more successful than intermediate learners in terms of fewer error counts. That is, they had fewer errors overall on both first and second drafts. These findings confirm the research results of Tseng (2014), who suggested that as writer proficiency increases, so does facility in correcting local errors.

We did not see evidence that language ability level affected writers' use of RA or TF. Indeed, students at both language ability levels generally performed better on their revised drafts because of TF as indicated above. The one exception was the error category of punctuation in which intermediate-level learners benefited more from TF than advanced-level learners. This is because advanced-level learners had fewer punctuation errors to begin with and were just as effective using TF as RA. In this one error category, advanced learners seemed to use TF and RA interchangeably the way Hartwell (1985) predicted. But in all other areas, TF was superior for both groups. So, with this finding in mind, it is perhaps justifiable to assert that when advanced L2 students read their own writing aloud, they can correct most of their *punctuation* errors by intuition.

Our study showed that with RA, students were able to correct some surface errors. However, RA was not as effective as TF, which led to even more corrected surface errors than RA did. This indicates that regardless of the classroom or course context, teachers and tutors should not neglect the importance of TF. We recommend that RA be a supplement to TF but not a replacement. Furthermore, composition teachers and writing center tutors may need to take the time to guide students through the RA process or model it in order for it to be maximally useful.

This study on student writing has various strengths and limitations that must be considered as well as avenues for future research. One of the strengths of this study was its ecological validity. University students usually produce short texts in class or during exams and revise their work quickly. Consistent with recent studies (Kim & Emeliyanova, 2021; McCarthy et al., 2022) we determined that 50% of composition time was a realistic amount of time for students to revise their work. Additionally, students often do not have an audience to read to when revising, so reading to oneself is an authentic scenario, though having participants read their work to another person may offer more validity and ensure task completion. Another strength of this study was the focus on grammar, as writers often concentrate on formal features when revising. Furthermore, we achieved high interrater reliability (> .85 Kappa) and double-rated every paragraph in the data coding process. Finally, the coding scheme used in this study was developed for L2 writers and has been employed in previous research studies (e.g., Eckstein & Ferris, 2018; Eckstein et al., 2020). However, future research might compensate for some of the limitations of this study by, for example, collecting data over a longer period of time, since students' RA skills may improve diachronically. Furthermore, the present study only examined international L2 writers in a non-matriculated intensive English program, so future RA research could explore its effects in additional contexts and with different learners, such as in K–12 or foreign language classes or writing centers and with immigrant writers and university students. It is also possible that writers at more extreme language ability levels would perform differently on RA tasks. Future research should examine beginning-level L2 writers as well as superior-level writers to broaden the spectrum and better reveal how language ability affects RA and TF proofreading.

This study brings some significant implications to writing tutors and composition teachers. While RA has already been heavily promoted in different writing contexts, this study showed that RA should not be used as a replacement to TF. Teacher feedback still holds an important position in error feedback processes, and independent study or online courses should not attempt to replace this valuable source of feedback with RA alone. Instead, RA can be used as a supplement, but first, it requires instructors to explicitly guide students through this process (Rowe, 2010). This might be done by modeling reading aloud to students on a draft with few errors and demonstrating the process of reconsidering the accuracy of a particular grammatical structure. Teachers can then model the process of turning to a grammar guide or dictionary for assistance in correcting the error. Teachers might then work with individual students, asking students to read aloud while the teacher monitors the reading and points out and discusses places where the oral production deviates from the text. Teachers might make use of an audio recording to help students hear those deviations in order to evaluate grammatical correctness. Again, all of this should be done as a supplement to TF. That is, students might read their writing aloud for language errors as a prerequisite to submitting it for teacher, peer, or tutor review. To what extent RA leads to a reduction in surface errors, instructors can better target errors that students cannot correct alone and otherwise focus on substantial areas for feedback.

About 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 Journal of Response to Writing. ORCID ID: 0000-0002-3667-4571

Ying Suet Michelle Lung is a PhD student in Linguistics at The University of Utah. She has experience teaching English language learners in Hong Kong, Japan, England, and the United States. Her passion extends beyond the classroom, as her current research interests focus on second language acquisition, writing development, and the role of feedback in learning. ORCID ID: 0000-0003-3167-3542

Natasha Gillette is an Assistant Professor of English Language Teaching and Learning at Brigham Young University-Hawaii. Her research interests include authentic language assessment, English for specific purposes and the academic success of indigenous student populations. ORCID ID: 0009-0001-9403-2147

To Cite this Article

Eckstein, G., Lung, Y.S.M., & Gillette, N. (2025). Can reading aloud replace teacher feedback? A comparative analysis of proofreading in ESL writing. *Teaching English as a Second Language Electronic Journal (TESL-EJ), 28*(4). https://doi.org/10.55593/ej.28112a5

References

ACTFL. (2012). ACTFL proficiency guidelines [Electronic version]. Retrieved October 22, 2021, from https://www.actfl.org/educator-resources/actfl-proficiency-guidelines

ACTFL. (2016). Assigning CEFR ratings to ACTFL assessments. https://www.actfl.org/uploads/files/general/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf

Azeez, P. Z. (2020). Investigating editing and proofreading strategies used by Koya University lecturers. *Journal of the University of Garmian, 7*(4). https://doi.org/10.24271/garmian.2070324

Bakla, A., & Karakaş, A. (2022). Technology and strategy use in academic writing: Native, native-like versus non-native speakers of English. *Ibérica, 44*, 285–314. https://doi.org/10.17398/2340-2784.44.285

Bartholomae, D. (1980). The study of error. *College Composition and Communication, 31*, 253–269. https://doi.org/10.2307/356486

Block, R. (2016). Disruptive design: An empirical study of reading aloud in the writing center. *The Writing Center Journal, 35*(2), 33-59. https://www.jstor.org/stable/43824056

Borrowman, S. (2004). The trauma of reading aloud. *Teaching English in the Two Year College, 32*(2), 197–198. https://www.proquest.com/docview/220941734?pq-origsite=gscholar&fromopenview=true

Çetinkaya, G. (2020). A Comparative evaluation on silent and read-aloud revisions of written drafts. *International Journal of Curriculum and Instruction*, *12*(2), 560-572. https://eric.ed.gov/?id=EJ1271126

Chartered Institute of Editing and Proofreading. (2020). *Ensuring editorial excellence: The CIEP code of practice.* https://www.ciep.uk/standards/code-of-practice/section-2.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Cogie, J., Strain, K., & Lorinskas, S. (1999). Avoiding the proofreading trap: The value of the error correction process. *The Writing Center Journal, 19*(2), 7-32. https://www.jstor.org/stable/43442834

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. H. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Wiley. https://doi.org/10.1002/9780470757642

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://psycnet.apa.org/fulltext/2001-16162-009.html

Cushing, C., & Bodner, G. E. (2022). Reading aloud improves proofreading (but using Sans Forgetica font does not). *Journal of Applied Research in Memory and Cognition, 11*(3), 427–436. https://doi.org/10.1037/mac0000011

Daneman, M., & Stainton, M. (1993). The generation effect in reading and proofreading. Is it easier or harder to detect errors in one's own writing? *Reading and Writing: An Interdisciplinary Journal, 5*, 297–313. https://doi.org/10.1007/BF01027393

Davis, J. R. (1995). Rewriting thoughts on proofreading. *Research and Teaching in Developmental Education, 11*(2), 85–91. https://www.jstor.org/stable/42801885

Eckstein, G., & Ferris, D. (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

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. https://doi.org/10.18806/tesl.v37i2.1339

Elbow, P. (2012). *Vernacular eloquence: What speech can bring to writing.* Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199782505.001.0001

Ferris, D. R. (2011). *Treatment of error in second language student writing* (2nd Ed.). University of Michigan Press.

Gibson, S. (2008). Reading aloud: A useful learning tool? *ELT journal*, *62*(1), 29–36. https://doi.org/10.1093/elt/ccm075

Hale, A. D., Skinner, C. H., Williams, J., Hawkins, R., Neddenriep, C. E., & Dizer, J. (2007). Comparing comprehension following silent and aloud reading across elementary and secondary students: Implication for curriculum-based measurement. *The Behavior Analyst Today, 8*(1), 9–23. https://doi.org/10.1037/h0100101

Hampton, A. R. (2019). *Proofreading power: Skills & drills.* Cornerstone Communications.

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. https://doi.org/10.5054/tq.2010.213781

Hartwell, P. (1985). Grammar, grammars, and the teaching of grammar. *College English*, *47*(2), 105–127. https://www.ou.edu/hartwell/Hartwell.pdf

Hedgcock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language writing instruction. *Journal of second language writing*, *1*(3), 255–276. https://doi.org/10.1016/1060-3743(92)90006-B

Kim, Y., & Emeliyanova, L. (2021). The effects of written corrective feedback on the accuracy of L2 writing: Comparing collaborative and individual revision behavior. *Language Teaching Research, 25*(2), 234–255. https://doi.org/10.1177/13621688198314

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*(1), 1–44. https://doi.org/10.1111/0023-8333.101997010-i1

Larigauderie, P., Guignouard, C. & Olive, T. (2020). Proofreading by students: Implications of executive and non-executive components of working memory in the detection of phonological, orthographical, and grammatical errors. *Read Writ 33*, 1015–1036. https://doi.org/10.1007/s11145-019-10011-6

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge. https://www.mdthinducollege.org/ebooks/statistics/A_Guide_to_Doing_Statistics_in_Second_Language_Research_Using_SPSS.pdf

Lott, D. A. (2016). *Reading your work aloud—a crucial step in your writing process.* Los Angeles Editors & Writers Group. https://laeditorsandwritersgroup.com/reading-work-aloud-crucial-step-writing-process/

Mackay, E., Lynch, E., Sorenson Duncan, T., & Deacon, S. H. (2021). Informing the science of reading: Students' awareness of sentence-level information is important for reading comprehension. *Reading Research Quarterly, 56*, S221–S230. **https://doi.org/10.1002/rrq.397**

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*(4), 390–395. https://doi.org/10.1177/0963721417691356

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 671–685. https://doi.org/10.1037/a0018785

Malin, N. L. (2019). *Efficacy of incorporating text-to-speech in the composition classroom.* (Publication No. 0000-0002-8706-2364) [Doctoral dissertation, Texas Woman's University]. Repository at TWU. https://www.proquest.com/docview/2600306462?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses

Matsuda, P. K., & Cox, M. (2009). Reading an ESL writer's text. In S. Bruce & B. Rafoth (Eds.), *ESL writers: A guide for writing center tutors* (2nd ed., pp. 42–50). Boynton/Cook Publisher. https://doi.org/10.37237/020102

McCarthy, K., Roscoe, R., Allen, L, Likens, A., & McNamara, D. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing, 52,* 1–21. https://doi.org/10.1016/j.asw.2022.100608

Norouzian, R., & Plonsky, L. (2018). Eta-and partial eta-squared in L2 research: A cautionary review and guide to more appropriate usage. *Second Language Research, 34*(2), 257–271. https://doi.org/10.1177/0267658316684904

Perl, S. (2006). The composing process of unskilled college writers. In V. Villanueva (Ed.), *Cross-talk in comp theory* (pp. 17–42). National Council of Teachers of English. https://openlab.citytech.cuny.edu/fywpd/files/2019/01/pearl-composing-processes.pdf

Pilotti, M., Chodorow, M., Agpawa, I., Krajniak, M., & Mahamane, S. (2012). Proofreading for word errors. *Perceptual and motor skills, 114*(2), 641–664. https://doi.org/10.2466/22.24.27.PMS.114.2.641-664

Pilotti, M., Maxwell, K., & Chodorow, M. (2006). Does the effect of familiarity on proofreading change with encoding task and time? *The Journal of General Psychology, 133*(3), 287–299. https://doi.org/10.3200/GENP.133.3.287-299

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Powers, J. K. (1993a). Bending the "rules": Diversifying the model conference for the ESL writer. *Writing Lab Newsletter, 17*(6), 1–8.

Powers, J. K. (1993b). Rethinking writing center conferencing strategies for the ESL writer. *The Writing Center Journal, 13*(2), 39–47. https://www.jstor.org/stable/43441929

Purcell, K. C. (1998). Making sense of the meaning: ESL and the writing center. *The Writing Center Journal, 22*(6), 1–4. https://wac.colostate.edu/docs/wln/v22/22.6.pdf

Rafoth, E. (2017). *Why we ask to read the paper aloud.* Writing Center at John Carroll University. https://jcuwritingcenter.wordpress.com/2017/02/17/why-we-ask-to-read-the-paper-aloud/

Riefer, D. M. (1993). Behavior engineering proposals: 5. An experimental comparison of team versus solo proofreading. *Perceptual and motor skills*, *76*(1), 111-117. https://doi.org/10.2466/pms.1993.76.1.111

Robinson, M. F., Meisinger, E. B., & Joyner, R. E. (2019). The influence of oral versus silent reading on reading comprehension in students with reading disabilities. *Learning Disability Quarterly, 42*(2), 105–116. https://doi.org/10.1177/0731948718806665

Rowe, D. (2010). *"What feels good in the mouth and sounds right to the ear": An examination of the practice of reading aloud during revision.* (Publication No. 3435858) [Doctoral dissertation, Rensselaer Polytechnic Institute]. ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/816596740?pq-origsite=gscholar&fromopenview=true

Ryan, L, & Zimmerelli, L. (2006). *The Bedford guide for writing tutors* (4th ed.). Bedford/St. Martins.

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition, 131*(1), 1–27. https://doi.org/10.1016/j.cognition.2013.11.018

Shafto, M. A. (2015). Proofreading in young and older adults: The effect of error category and comprehension difficulty. *International Journal of Environmental Research and Public Health, 12*(11), 14445–14460. https://doi.org/10.3390/ijerph121114445

St. John, R. L. (2004). *An analysis of the self-evaluation strategy of reading one's drafts aloud as an aid to revision: A multi-modal approach.* (Publication No. LD2489.Z68 2004 .S72) [Doctoral dissertation, Ball State University]. Ball State Theses and Dissertations. https://cardinalscholar.bsu.edu/items/0cdac195-bd18-4ca4-9239-9533ee3711fd

The Writing Center, University of North Carolina at Chapel Hill. (2024). *Reading aloud*. https://writingcenter.unc.edu/tips-and-tools/reading-aloud/

Tseng, T. J. L. (2014). The Role of Reading Aloud in EFL Writing Revision. *Journal of the College of Liberal Arts, National Changhua Normal University, 9*, 221-252. https://www.airitilibrary.com/Article/Detail/P20130308005-201403-201509170013-201509170013-221-252

Willis, J. (2008). *Teaching the brain to read: Strategies for improving fluency, vocabulary and comprehension.* Association for Supervision and Curriculum Development.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24, 1131–1161.* https://doi.org/10.1037//0096-1523.24.4.1131

Appendix A: Ns, Means, SDs, and Range for all Measures by Draft, Group, and Class (Proficiency level)

Measure	Draft	Group	Proficiency	N	M	SD	Range
Mechanics	1	TF	Intermediate	63	5.84	5.34	22.13
			Advanced	53	2.423	1.713	6.86
		RA	Intermediate	63	4.885	4.016	17.21
			Advanced	53	3.094	3.379	14.67
	2	TF	Intermediate	57	2.729	4.167	21.7
			Advanced	45	0.65	0.844	3.57
		RA	Intermediate	61	3.837	3.463	13.28
			Advanced	51	1.684	2.585	12.5
Nouns/noun phrases	1	TF	Intermediate	63	3.852	3.249	13.35
			Advanced	53	1.978	2.745	17.42
		RA	Intermediate	63	3.736	3.204	14.06
			Advanced	53	1.711	1.817	8.24
	2	TF	Intermediate	57	2.718	2.896	11.51
			Advanced	45	0.891	1.525	6.68
		RA	Intermediate	61	3.382	2.98	13.78
			Advanced	51	1.291	1.485	6.02
Pronoun usage	1	TF	Intermediate	63	0.254	0.653	2.87
			Advanced	53	0	0	0
		RA	Intermediate	63	0.239	0.552	1.9
			Advanced	53	0.264	0.494	2.2
	2	TF	Intermediate	57	0.221	0.576	2.14
			Advanced	45	0	0	0
		RA	Intermediate	61	0.207	0.524	1.93
			Advanced	51	0.203	0.464	2.25
Punctuation	1	TF	Intermediate	63	2.376	2.026	7.64
			Advanced	53	1.789	1.393	5.34
		RA	Intermediate	63	2.054	1.926	6.35
			Advanced	53	1.259	1.27	3.75
	2	TF	Intermediate	57	1.602	1.444	5.23
			Advanced	45	1.658	1.393	5.56
		RA	Intermediate	61	1.921	1.941	8.18
			Advanced	51	0.955	1.076	3.5
Sentence structure	1	TF	Intermediate	63	6.379	3.408	18.33
			Advanced	53	3.933	2.351	9.85
		RA	Intermediate	63	4.989	3.377	15.52
			Advanced	53	3.443	2.266	10.44
	2	TF	Intermediate	57	4.925	3.39	14.67
			Advanced	45	3.236	2.133	9.47
		RA	Intermediate	61	4.564	3.115	13.97
			Advanced	51	3.135	2.099	9.45

Subject-verb agreement	1	TF	Intermediate	63	1.095	1.446	5.09
			Advanced	53	0.4	0.653	3.05
		RA	Intermediate	63	0.779	1.153	4.31
			Advanced	53	0.437	0.691	2.24
	2	TF	Intermediate	57	0.638	1.183	4.82
			Advanced	45	0.118	0.332	1.55
		RA	Intermediate	61	0.79	1.166	5.63
			Advanced	51	0.323	0.607	2.7
Verbs/verb phrases	1	TF	Intermediate	63	0.891	1.846	7.85
			Advanced	53	0.515	0.74	3.4
		RA	Intermediate	63	2.311	2.275	8.44
			Advanced	53	0.653	0.946	3.24
	2	TF	Intermediate	57	0.54	1.329	5.79
			Advanced	45	0.293	0.531	1.8
		RA	Intermediate	61	1.877	1.741	7.67
			Advanced	51	0.513	0.784	2.77
Word form	1	TF	Intermediate	63	1.634	1.776	8.31
			Advanced	53	0.871	1.301	5.51
		RA	Intermediate	63	1.321	1.247	5.51
			Advanced	53	1.04	1.024	4.3
	2	TF	Intermediate	57	1.291	1.701	7.71
			Advanced	45	0.227	0.419	1.26
		RA	Intermediate	61	1.181	1.183	5.06
			Advanced	51	0.808	1.071	4.27
Word choice	1	TF	Intermediate	63	3.389	2.53	11.8
			Advanced	53	1.426	1.379	5.53
		RA	Intermediate	63	2.959	2.4	12.16
			Advanced	53	1.353	1.479	7.34
	2	TF	Intermediate	57	2.743	2.413	8.57
			Advanced	45	0.767	1.019	4.66
		RA	Intermediate	61	2.645	2.134	10.85
			Advanced	51	1.271	1.524	6.14
Total	1	TF	Intermediate	63	25.71	11.34	58
			Advanced	53	13.33	5.59	27.8
		RA	Intermediate	63	23.27	9.92	50.3
			Advanced	53	13.26	7.98	39.6
	2	TF	Intermediate	57	17.4	11.99	56
			Advanced	45	7.84	3.99	17.7
		RA	Intermediate	61	20.4	8.87	42.2
			Advanced	51	10.18	7.22	33.4

Appendix B: Statistical output for all main and interaction effects

Simple Main Effects

Measure	Draft			Group			Class		
	f	*p*	η^2	*f*	*p*	η^2	*f*	*p*	η^2
Mechanics	139.16	< .001	0.022	0.77	0.38	0.001	25.11	< .001	0.036
Nouns/noun phrases	43.23	< .001	0.007	0.12	0.727	0.000	32.84	< .001	0.047
Pronoun usage	3.29	0.071	0.001	2.83	0.094	0.004	3.48	0.063	0.005
Punctuation	17.69	< .001	0.003	1.63	0.203	0.002	7.20	0.008	0.011
Sentence structure	67.38	< .001	0.007	3.88	0.05	0.005	23.70	< .001	0.035
Subject-verb agreement	20.02	< .001	0.003	0.00	0.946	0.000	12.39	< .001	0.019
Verbs/verb phrases	19.69	< .001	0.002	15.70	< .001	0.023	21.50	< .001	0.031
Word form	22.57	< .001	0.006	0.25	0.618	0.000	13.50	< .001	0.02
Word choice	37.55	< .001	0.004	0.03	0.855	0.000	40.36	< .001	0.058
Total	306.39	< .001	0.023	0.11	0.741	0.000	74.30	< .001	0.101

Interaction Effects

Measure	Draft x Group			Draft x Class			Draft x Group x Class		
	f	*p*	η^2	*f*	*p*	η^2	*f*	*p*	η^2
Mechanics	18.90	<.001	0.003	1.61	0.205	0.000	3.52	0.062	0.001
Nouns/noun phrases	12.90	<.001	0.002	0.00	0.964	0.000	0.40	0.527	0.000
Pronoun usage	0.31	0.581	0.000	0.09	0.770	0.000	0.94	0.333	0.000
Punctuation	1.30	0.256	0.000	2.65	0.105	0.000	9.69	0.002	0.002
Sentence structure	24.83	<.001	0.003	7.18	0.008	0.001	4.74	0.031	0.001
Subject-verb agreement	16.78	<.001	0.003	0.89	0.348	0.000	1.78	0.183	0.000
Verbs/verb phrases	0.71	0.400	0.000	1.23	0.268	0.000	0.53	0.467	0.000
Word form	4.57	0.034	0.001	2.35	0.126	0.001	1.27	0.260	0.000
Word choice	11.97	<.001	0.001	0.07	0.793	0.000	0.55	0.461	0.000
Total	65.08	<.001	0.005	3.40	0.067	0.000	6.23	0.013	0.000