

The effects of teacher background on how teachers assess L1-like and L2-like grammar errors: An eye-tracking study

Grant ECKSTEIN, Brigham Young University, USA
Wesley Kaoto SCHRAMM, Brigham Young University, USA
Kate MATTHEWS, Brigham Young University, USA¹

Although composition teachers and L2 writing teachers may have similar objectives, researchers have shown that these groups of teachers may focus on different writing features and may even have differing views on error gravity when assessing student writing (Brown, 1991; Elder et al., 2003; Rifkin & Roberts, 1995). Common methods for examining these differences include analyzing rater scores and using reflective protocols. Only one study has used eye-tracking methodology to explore the raters' reading behaviors (Eckstein et al., 2018). The current study expounds on that study to examine whether L2 or composition teachers rate differently. Three L1-like errors and three L2-like errors were identified and introduced into eight paragraphs. Composition and L2 writing teachers assessed the eight paragraphs while an eye-tracker measured their eye-movements. Results indicated that L2 writing teachers assigned overall higher scores to L2 students than composition teachers. Although both composition teachers and L2 teachers may have similar teaching objectives, when rating L2 papers, different scores are assigned.

Keywords: Composition; Error Gravity; Eye-Tracking; Grammar; L2 Writing

1. Introduction

Compositionists and L2 writing teachers, although from different disciplines, both have similar goals: helping students become more effective academic writers in the English language. Composition teachers tend to teach students whose native language is English (L1 writers) whereas L2 writing teachers teach students whose first language is not English (L2 writers). The number of international students studying in US universities has been steadily increasing for several decades (Institute of International Education, 2017), making it more common for mixed composition classes with both L1 and L2 English speakers (Ferris, 2011; Harklau, 1994). In response to this phenomenon, researchers have investigated how composition teachers and L2 writing

¹ Corresponding Author (Email: grant_eckstein@byu.edu)

teachers assess L2 writing differently (Elder et al., 2003) and what features of writing both groups tend to focus on (Brown, 1991). Such research has shown that composition teachers can be harsher than L2 writing teachers, and that composition and L2 writing teachers often focus on different writing features.

Researchers investigating why composition and L2 writing teachers assess writing differently have examined areas such as teachers' attitudes towards L2 students (Ferris et al., 2011; Matsuda et al., 2013; Shvidko, 2015), rubrics used for assessment (Barkaoui, 2010; Lindsey & Crusan, 2011; Song & Caruso, 1996), experiences of teachers (Cumming, 1990; Santos, 1988; Song & Caruso, 1996), ethnolinguistic biases (Janopoulos, 1992; Lindsey & Crusan, 2011; Rubin & Williams-James, 1997), and views of error gravity (Brown, 1991; Elder et al., 2003; Santos, 1988; Song & Caruso, 1996; Sweedler-Brown, 1993). It is also possible that differences are manifestations of teachers' ideological differences within the fields of composition and L2 writing (Atkinson & Ramanathan, 1995; Eckstein et al., 2018; Santos, 1992). L1 and L2 writing may differ across features of rhetorical structure (Connor, 2011), cohesive devices and organizational patterns (Leki et al., 2008), lexical variation (Crossley & McNamara, 2009), and number and type of language errors (Eckstein & Ferris, 2018; Rifkin & Roberts, 1995). Yet, there has been little research examining teachers' reading behaviors as a potential source of difference (e.g., Eckstein et al., 2018; Eckstein et al., 2019). Eye-tracking is an emerging method in language research (Conklin & Pellicer-Sanchez, 2016) that records physical and temporal reading behavior, which may be analyzed for differences in what composition and L2 writing teachers look at when assessing writing.

2. Background

2.1. Differences between composition and L2 writing

Prior to what some called a "disciplinary division of labor", English composition studies subsumed L2 writing, and L2 students attended classes alongside L1 peers. However, shortly after World War II, an increase in international students in US universities led to a demand for teachers who were trained in teaching L2 writers (Ferris, 2009; Matsuda, 1999). Compositionists and applied linguists eventually agreed to work with the different groups respectively in the 1960s, which resulted in a professional and ideological divide with compositionists focused on rhetoric and creative writing and L2 writing teachers focused on language (Santos, 1992). The two fields also developed distinct cultural beliefs and ideologies. Atkinson and Ramanathan (1995), for instance, identified several ideological differences between an English composition and English language program within a single university. In their ethnographic comparison, the researchers found that teachers differed in their assumptions of students' cultural knowledge, the

metagoals of the programs, and the form and content of student writing. They found that composition teachers assumed that students shared Western cultural knowledge (e.g., shared understanding about what originality or critical thinking meant), and thought that instruction should go beyond students' academic needs and should encourage sophisticated communication rather than formulaic writing. On the other hand, L2 English instructors did not assume a shared Western cultural knowledge; they focused on practical skills that were useful for students' academic needs and encouraged clear and straightforward communication, often in the form of formulaic writing.

Other researchers, such as Santos (1992), have similarly observed disciplinary differences, arguing that composition encourages students to examine and challenge power structures that control their lives while L2 writing provides students with skills to meet the immediate writing requirements of their academic coursework without attempting to change their "sociopolitical consciousness" (p. 9). Costino and Hyon (2011), have further argued that the two fields use separate vocabulary (power, ideology, critical in composition and skills and practice in L2 writing). This, and the fact that the two fields attend separate conferences (e.g., CCCC and TESOL), publish in separate journals (e.g., CCC and TESOL Quarterly), and use separate referencing styles in publications (i.e., MLA and APA), reflects Ferris' (2009) lament that "there is still a discouraging lack of communication among composition professionals (L1 and L2) in post-secondary contexts" (p. 147).

Thus, although composition and L2 writing have similar origins and are ostensibly both designed to provide English writing instruction to college students, they nevertheless approach that goal from different epistemologies, employ different tactics, and communicate with different vocabularies. An important outcome of this divide is that teachers within these fields tend to view and assess student writing differently as well (Cumming et al., 2001; Eckstein et al., 2018; Eckstein et al., 2019). Cumming et al. (2001) found that ESL instructors evaluate student writing in a step-by-step process of decision making while compositionists tend to withhold judgement of an essay until completing it (Edgington, 2005; Wolfe, 2005). Moreover, Eckstein et al. (2018) showed that L2 writing teachers spent more time looking at student texts and also focused more on rhetorical features and less on grammatical features than did compositionists. This last point, about differences in the way teachers view grammar, is especially poignant given the different ways that the two disciplines tend to conceptualize it.

2.2. Composition teachers' view of grammar

Historically, compositionists have conceptualized the notion of grammar in a very complex and nuanced way. Hartwell (1985) who expands on the thinking

of Francis (1954), posits five meanings of the ambiguous term “grammar.” In list form, Hartwell explains them as (1) grammar in a writer’s head, (2) scientific grammar, (3) usage, (4) school grammar, and (5) stylistic grammar. The first grammar refers to tacit and intuitive knowledge about language that proficient L1 speakers share, including word order, noun pluralization, subject-verb agreement, and so forth. It is thought to be autonomous, impenetrable, and unalterable by explicit intervention (Chomsky, 1990). The second grammar refers to theoretical models, often designed by linguists, to account for the full system of language. Chomsky’s generative grammar or Bresnan’s lexical functional syntax (Bresnan et al., 2015) are examples of such models. Hartwell largely dismisses grammar three, that of usage, which represents formalized grammar rules about language, that fails to disrupt meaning within a text (Williams, 1981). Grammar four is perhaps the most recognized form of grammar—that of rules meant to describe how to produce “correct language,” such as adding an ‘s’ to form a plural noun. Finally, grammar five reflects handbook-style grammar which is meant to “improve the style” of already well-formed writing, such as when to use that rather than which.

Grammars three, four, and five are important for teachers who expect to teach L1 writers what can be done with English within the correct parameters of English rules. However, there has been some disagreement about what constitutes grammar errors within the composition paradigm. Historically, some viewed grammar errors broadly as non-standard syntax and morphology (Francis, 1954; Pelosi, 1973) while others focused on inappropriate usage, etiquette, or style (Francis, 1954; Kolln, 1981; Lance, 1977). More recently, composition researchers have empirically examined L1 English writer errors and determined that the most relevant grammar issues include punctuation, conjunctions, prepositions, confused words, and agreement problems (see Connors & Lunsford, 1988; Eckstein & Ferris, 2018; Gillespie & Lerner, 2000; Lunsford & Lunsford, 2008).

Even while teachers are aware of students’ grammar needs (Anson, 2000; Santa, 2006), there remains some resistance among composition instructors spending time on grammar instruction in writing classes (Ferris et al., 2017). This perspective may have originated in 1963 with the oft-quoted NCTE report by Braddock et al. that “the teaching of formal grammar has a negligible or . . . even a harmful effect on improvement in writing” (pp. 37-38). This statement has shaped composition dramatically (Kolln & Hancock, 2005) as have subsequent anti-grammar arguments, studies, and policies (Hartwell, 1985; Kolln & Hancock, 2005; Krashen, 1984; Truscott, 1996; Zamel, 1985). In fact, though some researchers since 1963 have been eager for more attention to grammar (e.g., McCleary, 1995; Rifkin & Roberts, 1995), the permeating attitude seems to be a dismissal of grammar instruction (Ferris et al., 2017;

Hartwell, 1985) even when control of grammar is expected of students (Harrington et al., 2001; Matsuda, 2012).

2.3. L2 writing teachers' views of grammar error

L2 writing teachers have approached formal grammar correction differently, and one reason may be that since they work with L2 students who lack the grammar intuition of L1 speakers, grammar correction is assumed to be helpful in developing students' competence (Ferris, 2009). Though grammar instruction has been the focus of some debate (e.g., Ferris, 1999; Truscott, 1996, 1999), it is nonetheless an accepted part of language acquisition pedagogy.

L2 writing has adopted a broader definition of "grammar," which focuses on language use and linguistic accuracy (Barkaoui, 2010; Eckstein & Ferris, 2018). Grammar errors can therefore be associated with word choice, spelling, punctuation, redundancy, and cohesion (Santos, 1988; Sweedler-Brown, 1993) as well as noun, verb, and sentence structure, word order, and article concerns (Eckstein & Ferris, 2018; Lane & Lange, 2012). Thus, L2 writing teachers may view "grammar errors" as a category that represents a large variety of language problems.

2.4. Differences in assessing writing

Differences in the way that grammar is conceptualized is thought to lead to differences in how grammar errors in writing affect those who read them. It is important to note that error is commonly defined as language issues that interfere with comprehension, that are not considered acceptable or normal, or that irritate or distract a reader (Rifkin & Roberts, 1995; Santos, 1988). Research in error gravity has investigated which error types affect comprehension the most or are most irritating for composition and/or L2 writing teachers (Brown, 1991; Elder et al., 2003; Santos 1988; Song & Caruso, 1996; Sweedler-Brown, 1993), differences between novice and experienced raters (Cumming, 1990; Santos, 1988; Song & Caruso, 1996), effects of ethnolinguistic biases in identifying errors (Janopoulos, 1992; Lindsey & Crusan, 2011; Rubin & Williams-James, 1997), and composition teachers' attitudes towards L2 students (Ferris, 2011; Matsuda et al., 2013; Shvidko, 2015).

In terms of error types, there is little consensus on which type of errors are more distracting or serious in readers' minds. Sheorey (1986) found that verb-related errors were most problematic while Leonard and Gilsdorf (1990) found sentence structure errors to be the most distracting. Tomiyana (1980) found article errors to be easy to correct and thus not substantially affect understanding. Burt (1975) distinguished global and local errors, the former

causing significant communication problems and disrupting sentence organization while the latter (e.g., inflections, articles, auxiliaries, quantifiers) affects a single sentence constituent. Lane and Lange (2012) suggest that a preponderance of local concerns can become a global issue. On the other hand, compositionists and writing center scholars have differentiated lower order concerns from higher order concerns, the former referring to surface mechanical errors generally, and the latter focused on non-language issues, such as content, idea development, and organization (Keh, 1990). Much error gravity research has demonstrated that raters of both groups tend to focus more on content issues above grammar errors when indicating problem areas in student writing, even when they are presented with grammatically inaccurate writing (Khalil, 1985; Santos, 1988).

Teacher experience also tends to mitigate some differences in the way teachers view grammar errors. Santos (1988) showed more experienced teachers showed less irritation towards grammar errors than less experienced teachers. Cumming (1990) asked novice and expert L2 writing teachers to assess 12 essays written by L2 students with different proficiency levels. The teachers' performance highlighted strategic differences between older and younger teacher including strategies to interpret and judge essay features with expert teachers using a wider variety of strategies in the assessment task. Song and Caruso (1996) also briefly covered teacher experience corroborating Santos' findings in that more experienced teachers tended to be more lenient than less experienced teachers, at least when holistic rubrics were used. These studies indicate that more experienced teachers use a wider variety of strategies to help them reach a decision, which may be more lenient on grammar errors made by L2 students.

In early ethnolinguistic bias research, researchers hypothesized that teachers may either find greater fault in L2 students' writing than there is (Land & Whitley, 1989) or make extra allowances because of the difficulties that these students face (Janopoulos, 1992). To test these hypotheses, Janopoulos (1992) asked teachers from a variety of disciplines to rate isolated sentences that each contained an error. Each sentence was guised as either that of an L1 or L2 writer to different raters. The results indicated that teachers from the humanities did not rate sentences differently based on perceived student ethnolinguistic background. Rubin and Williams-James (1997) conducted similar research in which composition teachers rated writing samples that were labeled as being written by a US student, a Danish student, or a Thai student. The writing samples had grammar errors inserted based on previous error gravity research. After the analysis, the composition teachers were shown to be more lenient toward what they perceived as L2 writing; in fact, those they believed to be Thai students received the highest scores, demonstrating a level of rater bias. Lindsey and Crusan (2011) repeated Rubin

and Williams-James' method but broadened the participants to include teachers from many disciplines and examined the scores these teachers assigned when using an analytic and a holistic rubric. Their results showed that guised L2 writing often received lower scores than L1 writing when scored analytically but higher scores when scored holistically. These studies show a wide range of ways teachers may react to L2 writing. Teachers may score more leniently or more harshly based on students' perceived ethnolinguistic background.

Researchers have also examined composition and L2 writing teachers' perceptions of the gravity of errors in L1 and L2 writing. In Santos' (1988) study, raters seemed to agree that lexical errors were the most serious while giving significantly higher ratings to language use than to content even though they considered on average the language use unacceptable. Brown (1991) asked eight professors in an English and ESL department to rate 112 compositions written by L1 and L2 students at the end of their First Year Composition (FYC) class. The findings showed that there were no significant differences in the scores assigned by both groups; however, a feature analysis showed that the composition teachers tended to focus more on cohesion and syntax while the L2 writing teachers attended to organization. In another study, Sweedler-Brown (1993) collected six L2 essays and prepared an original and a version with corrected sentence-level errors. Both essays were scored holistically by graders who had received no ESL training. The results showed correlation between sentence-level features and grammar/mechanics with overall score, but no correlation between rhetorical and organizational features with overall score, indicating that sentence-level grammatical errors affected the overall scores of L2 writing. A similar study (Song & Caruso, 1996) contradicted Sweedler-Brown's in that composition teachers seemed to give greater weight to content and rhetorical features than they did to language use errors. Finally, when investigating how composition teachers responded to disciplinary writing, Weigle et al. (2003) found that composition teachers seemed to score the writing harsher than L2 writing teachers, and that composition teachers focused most on content and grammar while L2 writing teachers focused on a wider variety of features. The results of these five studies show little agreement. Areas of contradiction include whether teachers are more lenient or harsh because of grammar errors, and if errors should even be a focus in the first place. Because of this, and because essay rating may have a large impact on students' grades, more research is needed to understand how composition and L2 writing teachers interact with grammar errors in student writing.

Because of their different epistemologies and views of grammar, compositionists and L2 writing teachers are thought to view student writing differently. Examining their provision of feedback to L2 students, Ferris et al.

(2011) surveyed composition and L2 writing teachers at eight sites. Teachers indicated greater confusion on how they should respond to L2 writing. Some teachers expressed negative attitudes towards L2 students and little concern for their challenges with grammar. Others expressed a desire to help but registered a lack of knowledge of how to help. Many of the teachers reported that they changed their approach and focused more on grammar than they normally would. Other survey studies have shown similar results (Ferris et al., 2015; Matsuda, 2013), suggesting that composition teachers especially struggle with the language needs of L2 writers.

Empirical investigations into the kinds of errors L1 and L2 students make have yielded important distinctions that may account for the difficulty in working with L2 writers. Ferris (2006), who examined over 5,000 errors in L2 student writing, found that the type and frequency of L2 errors differed compared to L1 written errors (see also Connors & Lunsford, 1988; Lunsford & Lunsford, 2008). Silva (1993) reported greater L2 error counts among verb, preposition, article, and noun categories, and Doolan and Miller (2012) found that L2 immigrant writers' errors included verb forms, prepositional phrases, and word forms. Eckstein and Ferris (2018) found that L2 students made significantly more errors than L1 writers in the form of verbs, nouns, sentence structure, word choice, and word form. L2-specific errors may impact the readability of L2 writing, particularly if readers are more familiar with typical L1 errors.

2.5. Eye-tracking

The methods that researchers have employed in order to examine teachers' interaction with grammar errors can be grouped into two general methods and a third more unusual approach. The first uses assessment tasks and direct scoring (e.g., Brown, 1991; Lindsey & Crusan, 2011) or teacher comments (e.g., Cumming, 1990; Rubin & Williams-James, 1997) to examine teachers' decision making. The second involves think-aloud or reflective protocols associated with an assessment task (e.g., Barkaoui, 2010; Song & Caruso, 1996). Although task/direct scoring and think-aloud studies have been informative, they are rather limited in their generalizability. For example, the implications that can be drawn from scores assigned to writing samples cannot be verified, and there is often a mismatch between what teachers report they think and do and what they actually do (Montgomery & Baker, 2007).

An additional approach is that of eye movement, which is increasingly used in language research (Conklin & Pellicer-Sanchez, 2016; Rayner, 1998). This may be due to its more direct measure of reading behavior that reduces reactivity (Paulson et al., 2007). Eye-tracking provides empirical data by measuring fixations (pauses in eye-movement when the eyes gather information) and

saccades (quick eye-movement between fixations) (Huey, 1908/1968; Just & Carpenter, 1987). Fixations are believed to reflect attentional focus, which in turn is thought to correlate with cognition, according to the eye-mind hypothesis (Conklin & Pellicer-Sancez, 2016; Rayner, 1997, 1998). Put simply, what one looks at is thought to be what one is considering. This is supported by evidence that readers fixate on problem areas and areas that are hard to understand more frequently and for longer durations than other areas (Frazier & Rayner, 1982). Applied to error assessment, the eye-mind hypothesis predicts that teachers will fixate longer or return to areas of grammar error more than well-formed text.

Eye movement data is exceptionally rich, and many measures can be classified into early or late reading measurements. Early measures are thought to reflect automatic word recognition and lexical access while late measures are thought to show strategic comprehension processes (Conkin et al., 2018). That is, when first encountering a word, readers may recognize letters, decode words, and make general sense of word meanings. Readers may return to a word to integrate and comprehend text that was not initially processed or to resolve meaning confusion.

A recent eye-tracking study (Eckstein et al., 2018) shows promise for enriching error gravity research. The researchers identified reading differences between composition and L2 writing teachers. The researchers asked five composition and five L2 writing teachers to read a single essay written by an L2 student and give a holistic score based on a rubric that was divided into four categories: rhetoric, organization, word choice, and grammar. Eye movement data showed that composition teachers spent more time on early reading measures of phrases containing grammar errors while L2 writing teachers skipped them more often in early reading and then spent more time in later reading, suggesting that composition teachers examined grammar errors when first encountering them while L2 writing teachers skipped them initially but then worked to comprehend the text by refocusing on the errors. However, because it was a preliminary study and because of the small participant pool and single essay stimuli, the results are difficult to generalize. Further, because the researchers did not define grammar errors, there was little consistency in their stimuli. The researchers also only employed one writing sample, increasing the chance that findings were related to the idiosyncratic nature of one writing sample rather than more representative L2 writing. Additional research is needed to gain a better understanding of how teachers assess grammar, and the current study builds upon the preliminary study in an attempt to address these questions:

1. To what extent do composition and L2 writing teachers rate L1 and L2 student writing differently?

2. To what extent do eye movements of composition teachers differ when assessing grammar errors in L1 and L2 writing?
3. To what extent do eye movements of L2 writing teachers differ when assessing grammar errors in L1 and L2 writing?
4. To what extent do eye movements of composition and L2 writing teachers compare when assessing grammar errors in L2 writing?

3. Method

3.1. Participants

A total of 29 teachers from a large university in the Western United States participated in the study; 15 were composition teachers, and 14 were L2 writing teachers. All were native speakers of English. The average age was 33 years old and the average experience was 7 semesters of teaching writing. Among the composition teachers, the average age was 34 years old (10 female), and they had on average 11 semesters of writing instruction experience. All had graduated with or were currently pursuing a master's degree in English. They received and attended departmental training at the start of every semester and attending weekly workshops with a group of fellow teachers to discuss any concerns related to their class. The L2 writing teachers were 33 years old on average (13 female) and had an average of 4 semesters of teaching experience. Three were experienced teachers who were teaching while also pursuing a graduate-level TESOL certificate. The rest had or were pursuing a master's degree in TESOL. The L2 writing teachers reported receiving two hours of writing instruction training at the start of each semester that they taught a writing class. They also received training and calibration on writing assessment at least once per semester.

3.2. Passages

We collected timed student essays which were written in a first-year composition class during the first week of the semester by students not assigned to the teachers in this study. The prompt asked students to describe their writing process in 500-700 words after they had reflected on a past writing experience. From these essays, we selected the introductory paragraphs from eight essays, four written by L1 writers and four by L2 writers. Paragraphs were not altered in content or organization; however, we introduced two types of additional grammar errors—those which are statistically more likely to be made by monolingual English speakers (L1-type), and those statistically more likely to be made by speakers of English as a second language (L2-type). One reviewer of a previous version of this manuscript questioned whether errors can be classified based on L1-type or

L2-type. While we acknowledge this distinction is necessarily rough and unlikely to perfectly reflect all language users, researchers such as Eckstein and Ferris (2018) and Eckstein and Chang (2022) have found statistically significant differences in the kinds of errors made by L1 and L2 writers. For instance, Eckstein and Ferris (2018) found that L2 writers made significantly more errors than L1 writers in the categories of verb, noun, sentence structure, word choice, and word form errors. Eckstein and Chang (2022) found similar results except that pronoun usage and subject-verb agreement also discriminated L1 from L2 writers though word choice did not. These results indicate that some errors are more likely to reflect L2-type errors than errors made by monolingual writers.

Table 1
Top 10 Most Common L1- and L2-Type Errors

L1-type Errors (Lunsford & Lunsford, 2008)	L2-type Errors (Company, 2012)
Wrong word	Spelling
Missing Comma	Word choice
Incomplete documentation	Determiner
Vague pronoun reference	Preposition
Spelling	Singular/plural forms
Quotation	Word form
Unnecessary comma	Punctuation
Capitalization	Subject-verb agreement
Missing word	Verb form
Faulty sentence structure	Verb tense

Note: Bolded words in this table refer to errors that are more specific to composition or L2 writers

Error type for this study was selected by referring to Lunsford and Lunsford's (2008) analysis of the 20 most common errors made by composition students, and Company's (2012) analysis of the 15 most common error types made in her study of L2 writers. The three most common error types appearing on each list but not on the other were selected for inclusion in order to reflect typical L1 and L2 errors (see bolded words in Table 1). The following are examples of sentences with each type of error:

- **Missing comma:** *Prior to starting an essay I fill my head with so many different ideas.*

- **Vague pronoun reference:** *By the end of it, the reader might not even remember what the thesis of the essay was and clueless as to what message I was trying to convey.*
- **Capitalization:** *The best words to describe myself as a writer would be "Reluctant" and "insecure."*
- **Determiner:** *I struggle tremendously with organizing my thoughts and displaying them into the words.*
- **Preposition:** *This realist notion is very apparent on my past essay.*
- **Singular/plural:** *I confuse many word and have a hard time remembering simple grammar rules.*

Table 2
Number of Errors Assigned to Each Paragraph

Paragraph	Cap	Det	Com	Prep	S/P	Pro	Total
1	0	2	1	1	2	3	9
2	1	1	1	1	2	3	9
3	3	2	2	0	1	1	9
4	2	1	0	2	2	2	9
5	2	2	1	2	1	1	9
6	0	3	2	2	1	1	9
7	3	0	2	2	2	0	9
8	1	1	3	2	1	1	9
Total	12	12	12	12	12	12	72

In order to verify that each error was indeed an error, the researchers asked a group of L2 writing teachers to read each paragraph and identify errors they found. Any intentional error that was not identified during this activity was revised. We then asked a linguistics professor and a composition professor to review the errors and state whether they considered each to be an error that would affect their judgement of the quality of the text. The professors found only a few errors to be insufficiently salient, and in these cases, we moved the error to a new location within the paragraph where it appeared more salient until all errors were easy to identify and clearly reflected violations of standard English grammar expectations. Table 2 (above) shows how the errors were dispersed throughout the eight paragraphs.

3.3. Rubric

The rubric (included in Appendix B) consists of four categories: content, organization, vocabulary, and grammar. Each category could be scored from 0 to 7. The rubric was adapted from the Connor-Linton and Polio's (2014), which is itself an adaptation of Jacobs et al.'s (1981) rubric. It was adapted for this study to refer to paragraphs instead of full essays and by combining language use and mechanics into one section labeled as grammar to better reflect the focus of this study.

3.4. Apparatus

The machine used in this study was an SR Research EyeLink 1000 Plus (spatial resolution of 0.01°) which sampled at 1000 Hz. This eye-tracker required participants to rest their head in a mounted headrest to ensure accurate measurements. A computer screen with a display resolution of 1600 x 900 (approximately 3.5 characters subtended 1° of visual angle) displayed the paragraphs and rubric and was positioned 63 centimeters from the participants. Paragraphs were double spaced in 16-point monospace Courier font and displayed with 1-inch margins on all sides.

3.5. Areas of interest

Each paragraph was coded into areas of interest (AOIs) for later analysis. AOIs are semantically meaningful sections from which eye-tracking software takes measurements (Conklin & Pellicer-Sancez, 2016). We coded each target error in the eight paragraphs as an AOI by selecting the word in which the error was found. In the case of a comma, the area of interest included the word preceding the missing comma.

3.6. Measurements

The following reading measurements (Table 3) were selected to answer the research questions. Definitions are adapted from Conklin et al.'s (2018) eye-tracking manual for second language research.

Researchers have distinguished early and late reading measures (Inhoff, 1984; Staub & Rayner, 2007) in which, according to Conklin & Pellicer-Sanchez (2016), "early measures tap into automatic processes and the initial stages of processing" (p. 455), which include letter recognition, lexical access, and text decoding. Later reading measures represent "strategic processing and include revisits and reanalysis that result from processing difficulty" and therefore include processes associated with comprehension and integration. We included measures associated both with early and late reading processes to

determine how readers might automatically decode or process text and then subsequently comprehend and integrate it.

Table 3
Eye-Tracking Measures Used in This Study

Early reading measures		
Measure	Definition	Hypothesized process
Skip count	An indication of whether the word was fixated upon during the first read.	estimation of text predictability, skimming
First fixation duration	The duration of the first fixation on the word.	decoding, word recognition
First run dwell time	The total duration of all fixations on the word before exiting the word the first time.	word recognition, general understanding of text
Late reading measures		
Measure	Definition	Hypothesized process
Second run dwell time	The total duration of all fixations on the word after it has been exited the first time.	word integration, syntactic processing
Total dwell time	The total duration of all fixations on the word during a trial.	late word processing, syntactic processing
Regression-in count	The number of times the reader returned to the word from the right.	confusion, syntax ambiguity

3.6. Procedure

At the beginning of each session, the participant completed a 9-point calibration and validation while seated at the eye-tracker with head fixed in a chinrest. Calibration and validation were repeated regularly to ensure accurate measurements throughout the session. Participants viewed a practice paragraph and were instructed to read the paragraph and prepare to score it while the paragraph was displayed on the computer screen. Under the participant's control, the rubric was displayed, and participants were instructed to verbally give the scores for each of the four subcategories as well as an overall score. They were not able to return to the paragraph once the rubric was displayed. The same procedure was followed for the remaining eight paragraphs which were randomized in a counterbalanced design among

the participants. The participants were not informed whether the writing was that of an L1 or L2 student.

3.7. Data analysis

Participants' reported rubric scores for each category (content, organization, vocabulary, grammar, and overall) were analyzed descriptively followed by a mixed-effects analysis comparing composition and L2 writing teachers' scores. Eye-tracking data were collected for L1 and L2 grammar errors and likewise subject to a mixed-effects analysis for each reading measure. The fixed effects were error category (L1/L2 error) and error type, and the random effects were rater and paragraph in which the error occurred. The natural log of the dwell time measurements (first fixation duration, first run dwell time, second fixation duration, second run dwell time, and total dwell time) were used in order to normalize the data so the assumptions of a mixed-effects analysis would be met (Whelan, 2008).

4. Results

Results are organized by research question and show rater scores first followed by data from eye-movement behavior. Generally, the results show differences and assigned scores, but little difference in how teachers from different classroom backgrounds visually attend to L1 or L2 grammar errors.

4.1. Differences in rubric scores

The *overall* category was an average of scores on content, organization, vocabulary, and grammar (see Table 4).

Table 4

Average Scores by Category Given by Composition and L2 Writing Teachers

Category	Teacher Type	<i>n</i>	<i>M</i>	Mean diff.	<i>df</i>	<i>p</i>
Content	Composition	15	4.59	-0.224	40.6	0.424
	L2 writing	15	4.81			
Organization	Composition	15	4.41	-0.327	40.6	0.246
	L2 writing	15	4.73			
Vocabulary	Composition	15	4.18	-0.632	40.6	0.028
	L2 writing	15	4.81			
Grammar	Composition	15	3.68	-1.016	40.6	0.001
	L2 writing	15	4.70			
Overall	Composition	15	4.22	-0.555	40.6	0.052
	L2 writing	15	4.78			

L2 writing teachers gave higher overall scores on student writing by about half a point, and while this did not reach statistical significance, the probability level ($p = .052$) is very close to significant and suggests a trend worth considering. Of the individual categories, grammar showed the largest mean difference in assigned scores (-1.016) where L2 writing teachers ($M = 4.7$) scored grammar a full point higher than that of composition teachers ($M = 3.7$). This was a significant effect ($p < .001$), as was that of the vocabulary category ($p = .028$).

4.2. Eye-tracking measurements of L2 grammar errors

A mixed-effects analysis for reading measures of L2 errors showed that no measures significantly distinguished composition and L2 writing teachers (see Table 5). Both groups skipped about 50% of L2 errors during first pass reading and then spent just under 250 ms on average viewing L2 errors that were not skipped during first run reading. Though not significant ($p = .08$), total dwell time showed the largest difference between the groups: L2 writing teachers spent 66.5 milliseconds longer on L2 errors than composition teachers.

Table 5
Comparison of Composition and L2 Writing Teachers When Looking at L2 Errors

	Teacher Type	Mean ms or count	Mean Diff.	<i>df</i>	<i>p</i>
Skip Count	Composition	0.54	-0.09	27.14	0.14
	L2 writing	0.63			
First Fixation Duration	Composition	205.6	-1.24	26.42	0.92
	L2 writing	206.9			
First Run Dwell Time	Composition	223.2	-7.95	25.85	0.54
	L2 writing	231.1			
Second Run Dwell Time	Composition	242.3	7.86	32.13	0.64
	L2 writing	234.4			
Total Dwell Time	Composition	373.2	-66.50	27.17	0.08
	L2 writing	439.7			
Regression-in Count	Composition	0.44	-0.15	27.18	0.15
	L2 writing	0.59			

4.3. Composition teachers' reading behavior

Results were slightly more meaningful when comparing composition teachers' reading of L1 and L2 errors. First fixation duration ($p = .034$) and first run dwell time ($p = .001$) both revealed significant differences, indicating a longer early reading time for L2 errors. Regression-in count was also close to significant ($p = .09$), and descriptively shows that composition teachers regressed to L2 errors more often than L1 errors.

Table 6

Composition Teachers' Reading Behavior of L1 and L2 Grammar Errors

	Error Type	Mean ms or count	Mean Diff.	df	p
Skip Count	L1	0.55	0.00	1015.3	0.93
	L2	0.55			
First Fixation Duration	L1	218.5	13.55	776.8	0.03
	L2	205.0			
First Run Dwell Time	L1	250.6	27.90	783.9	0.00
	L2	222.7			
Second Run Dwell Time	L1	231.1	-11.12	400.9	0.31
	L2	242.3			
Total Dwell Time	L1	397.8	22.04	787.3	0.24
	L2	375.8			
Regression-in Count	L1	0.37	-0.08	1034.2	0.09
	L2	0.45			

4.4. L2 writing teachers' reading behaviors

When examining L2 writing teachers' reading of L1 and L2 errors versus composition teachers reading of L1 and L2 errors, results were slightly different.

Table 7

L2 Writing Teachers' Reading Behavior of L1 and L2 Grammar Errors

	Error Type	Mean ms or count	Mean Diff.	df	p
Skip Count	L1	0.55	-0.08	990.1	0.01
	L2	0.63			
First Fixation Duration	L1	215.3	8.23	757.0	0.23
	L2	207.1			
First Run Dwell Time	L1	246.9	15.31	762.5	0.09
	L2	231.6			
Second Run Dwell Time	L1	238.9	3.79	430.7	0.76
	L2	235.1			
Total Dwell Time	L1	419.1	-22.37	761.4	0.30
	L2	441.4			
Regression-in Count	L1	0.41	-0.17	982.8	0.00
	L2	0.59			

Skip count ($p = .01$) showed that L2 teachers were less likely to skip L1 errors, and another measure of early reading, first run dwell time ($p = .09$), was also

close to significant and similarly showed more attention to L1 errors. Late reading measures showed further differences; regression-in count ($p < .001$) was significantly higher for L2 errors, which L2 writing teachers initially skipped more, indicating a greater need for teachers to revisit or reconsider L2 grammar errors.

5. Discussion

5.1. Rubric scores

Rubric scores assigned in this study were consistently lower among composition teachers, though the differences were non-significant for content and organization. Brown (1991) and Eckstein et al. (2018) both found non-significant results when comparing holistic teacher scores. However, in the present study, we did observe significant differences in the holistic scores for vocabulary and grammar, which corroborates results from researchers who have shown that composition teachers give lower scores than L2 writing teachers in general (Weigle et al., 2003) and especially in grammar (Sweedler-Brown, 1993; Weigle et al., 2003). The L2 writing teachers scored vocabulary higher than composition teachers, possibly because L2 writing teachers are accustomed to L2 writing, which tends to employ a smaller vocabulary than L1 writing (Crossley & McNamara, 2009). The grammar score differences, however, are more relevant to the purpose of this study, and it is likely that composition teachers were more effected by grammar errors or considered them graver while L2 writing teachers may have overlooked grammar errors in order to compensate for language development or because of their general familiarity with language errors (e.g., Eckstein et al., 2018; Eckstein et al., 2019). Sweedler-Brown (1993) further observed that composition teachers are harsh on grammar errors even when other features are strong. Beyond grammar and vocabulary, overall scores trended toward significance and indicate descriptively that composition teachers were harsher graders on average by about half a point.

Together these findings demonstrate that composition and L2 writing teachers tend to differ in their assessment of student writing, particularly in their judgement of grammar errors. In this sense, composition teachers appear to view grammar errors as more grave compared to L2 writing teachers. While these ratings show a clear difference, eye-tracking measures can help determine whether attentional focus precipitated these differing scores.

5.2. Eye movements and L2 grammar errors

A mixed-effects analysis found no statistically significant differences between the reading behaviors of composition and L2 writing teachers when viewing

L2 grammar errors. This was surprising because it contradicted the findings of Eckstein et al. (2018), which showed that composition teachers spent significantly more time during first run reading on grammar errors but L2 writing teachers returned to these errors significantly more in later reading. The present study further shows that total dwell time approached significance ($p = .08$) and descriptively suggests that L2 writing teachers spent more time overall attending to L2 errors, or at least considering them, than composition teachers.

Based on differing rubric scores, we expected that composition and L2 writing teachers would read and interact with errors differently. However, this does not seem to be the case; only total dwell time was close to significant. Thus we argue that reading behaviors do not precipitate differences in scoring; that is, teachers do not score errors more harshly because of the attentional focus they put on those errors. Instead, different attitudes, biases, or backgrounds at a cognitive level may lead to different scores. Teachers may emphasize different features of writing they think are most important, or they could implement different strategies for interpreting and judging a text (Cumming, 1990). They may interpret texts as more L1- or L2-like and modify scores in response to their perceptions of the writer's ethnolinguistic background (Lindsey & Crusan, 2011). Teachers may also place different weight on error types based on personal perceptions of what is a severe error or what is distracting for them (Hartwell, 1985). For instance, they may perceive L1 errors more or less leniently than L2 errors, which motivated our final analyses.

5.3. Composition teachers reading L1 and L2 errors

A mixed-effects analysis showed statistically significant differences between first run dwell time and first fixation duration, both early reading measures. Composition teachers initially attended longer to L1 errors than to L2 errors, suggesting longer recognition and decoding time. This corroborates findings by Eckstein et al. (2018) who likewise found that composition teachers had longer early reading measures for grammar errors, though their study did not distinguish L1 and L2 errors. Perhaps because composition teachers are attuned to L1 errors, they mentally note and possibly correct them when first reading. Eckstein et al. reported that two of their five composition raters focused on errors when they affected the flow of the argument. Composition teachers descriptively returned to L1 errors less frequently than L2 errors, supporting the speculation that composition teachers processed L1 errors in situ without returning to them later.

The greater attention to L1 errors, at least during early reading, suggests that composition teachers may assign more gravity to L1 errors. They may find such errors distracting to an argument or more difficult to parse initially. This could

be because missing commas and vague pronouns are included as L1 errors. Missing commas and vague pronouns tend to cause ambiguity on the sentence level rather than on the word level, which is why they may have taken longer to decode and resulted in greater distraction.

5.4. L2 writing teachers reading L1 and L2 errors

L2 writing teachers behaved in somewhat similar ways to composition teachers in that they gave more attention initially to L1 errors, but unlike composition teachers, they returned to L2 errors to a significant degree. L2 teachers initially skipped L2 errors more frequently than L1 errors, suggesting that the L2 errors did not initially impede meaning making or that the L1 errors required greater early processing. The skip count may also be due to L2 writing teachers' familiarity with the errors, which they may have strategically chosen to skip. Alternately, these errors may have been highly predictable and thus easy to overlook given the teachers' language instruction background. L2 writing teachers returned to the L2 errors, however, which we speculate is a matter of training in that L2 writing teachers may look back to confirm the correctness or incorrectness of a phrase.

In terms of error gravity, L2 writing teachers seemed less troubled by grammar errors overall, though their tendency to skip fewer L1 errors initially leads us to postulate that L1 errors are more distracting or harder to process. It may be that L2 errors are likewise distracting given that composition teachers returned to them, but that they failed to inhibit comprehension.

6. Conclusion

Our overall results showed that composition teachers assigned significantly lower scores to grammar than L2 writing teachers. This difference might have predicted that teachers from different disciplinary backgrounds read L1 and L2 student grammar errors differently. Results from eye-movement analysis, however, showed that composition and L2 writing teachers did not behave differently when viewing L2 language errors, though descriptively, L2 writing teachers spent longer overall looking at L2 errors compared to composition teachers. Conversely, there was evidence that composition and L2 writing teachers visually interacted with L1 and L2 errors differently. Composition teachers tended to give more attention during early reading processes to L1 errors while L2 writing teachers tended to overlook L2 errors initially but then regressed back to those errors in a later reading process. These behaviors do not necessarily contradict one another; in both cases, L1 errors seemed to demand more early reading attention than L2 errors for both groups of instructors. This observation leads us to conclude that L1 errors are more cognitively demanding than L2 errors and as such may account at least

partially for lower grammar scores among composition teachers who are less likely to be familiar with such errors.

A possible extension of this interpretation is that less familiar errors will result in more cognitive attention and thus lower scores. Future research could test this prediction by including errors not within the top ten most frequent L1 and L2 errors as this study did. Additionally, the connection between cognitive difficulty and rater scores is especially tenuous, and further research could explore this connection by introducing errors that require greater cognitive effort to resolve, such as inverted word order, significant spelling errors, or awkward expressions. In terms of gravity measures, this study purposefully sought to determine behavioral differences in eye-movements, but without raters' self-evaluation of the gravity of each error, it is difficult to make inferences about whether one error was interpreted as more grave than another. Future research should thus include a survey component wherein raters can express their perception of each error's gravity.

The present study is limited in a number of ways that reduces its generalizability. The inclusion of six errors that are mostly mechanical in nature makes it difficult to generalize these findings to broader categories of grammar errors. Further, while the errors selected were intended to represent typical L1 and L2 errors, and some control was exerted to ensure the distinction, the reality is that all writers can potentially make all errors, so a tidy distinction between L1 and L2 errors is more theoretical than practical. The raters in this study were also all trained composition or writing teachers, and their experiences and backgrounds may differ greatly from disciplinary instructors. Finally, the eye-tracking procedure itself was a limiting factor in that the paragraphs and the rubric were always displayed separately due to screen space limitations. However, teachers may prefer to see rubrics and writing side-by-side in order to move back and forth throughout the assessment task. This limitation could have impacted late reading measurements, which may become more salient after referring to a rubric. Further research may show that there are differences in late reading measures when this limitation is taken into consideration. One possible solution may be to use a TOBII tracker apparatus, which does not require a headrest so that participants can have a paper rubric to consult.

Ultimately, the findings from this study have implications for both students and teachers. Students may be heartened or distressed to learn about the attention required by teachers to process their grammar errors. L1 students in particular can benefit from knowing this as part of an in-class discussion on the merits of editing and proofreading. If students can reduce the errors in their writing, it may have an impact on the cognitive work required to understand the text which could have a grade impact. L2 students may be heartened to know that

typical L2 errors are not as cognitively demanding as typical L1 errors, and when L2 writing teachers view these errors, they may be especially lenient during grading. Understanding the behavioral responses to L1 and L2 grammar errors can therefore have an important impact on L1 and L2 student writing.

Authors' Statement

The first two authors discussed and conceived the article together. Kate Matthews contributed to background research and overall idea development. The authors would like to thank Don Chapman and Dave Eddington for critical assistance during the early development of this research.

The Authors

Grant Eckstein (Email: Grant_eckstein@byu.edu) 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*.

Wesley Schramm (Email: wschramm4649@gmail.com) is a graduate of BYU's TESOL MA program. He teaches English classes in Japan. His research interests include language pedagogy and assessment.

Kate Matthews (Email: hadfielk@byu.edu) is a graduate student of BYU's TESOL MA program. She teaches English classes in Utah. Her research interests include second language acquisition, with an emphasis on feedback.

References

- Anson, C. M. (2000). Response and the social construction of error. *Assessing Writing*, 7(1), 5-21. [https://doi.org/10.1016/S1075-2935\(00\)00015-5](https://doi.org/10.1016/S1075-2935(00)00015-5)
- Atkinson, D., & Ramanathan, V. (1995). Cultures of writing: An ethnographic comparison of L1 and L2 university writing/language programs. *TESOL Quarterly*, 29(3), 539-568. <https://doi.org/10.2307/3588074>
- Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7, 54-74. <https://doi.org/10.1080/15434300903464418>
- Braddock, R., Lloyd-Jones, R., & Schoer, L. (1963). *Research in written communication*. National Council of Teachers of English.

- Bresnan, J., Asudeh, A., Toivonen, I., & Wechsler, S. (2015). *Lexical-functional syntax*. John Wiley & Sons.
- Brown, J. D. (1991). Do English and ESL faculties rate writing samples differently? *TESOL Quarterly*, 25(4), 587-603. <https://doi.org/10.2307/3587078>
- Burt, M. K. (1975). Error analysis in the adult EFL classroom. *TESOL Quarterly*, 9(1), 53-63. <https://www.doi.org/10.2307/3586012>
- Chomsky, N. (1990). Language and mind. In D. H. Mellor (Ed.), *Ways of communicating* (pp. 56-60). Cambridge University Press.
- Company, M. (2012). Error frequencies among ESL writers: A resource guide. *All Theses and Dissertations*, Paper 3420.
- Conklin, K., & Pellicer-Sanchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, 32(3), 453-467. <https://doi.org/10.1177/0267658316637401>
- Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.
- Connor, U. (2011). *Intercultural rhetoric in the writing classroom*. University of Michigan Press.
- Connor-Linton, J., & Polio, C. (2014). Comparing perspectives on L2 writing: Multiple analyses of a common corpus. *Journal of Second Language Writing*, 26, 1-9. <https://doi.org/10.1016/j.jslw.2014.09.002>
- Connors, R., & Lunsford, A. (1988). Frequency of formal errors in current college writing, or Ma and Pa Kettle do research. *College Composition and Communication*, 39(4), 395-409. <https://doi.org/10.2307/357695>
- Costino, K., & Hyon, S. (2011). Sidestepping our “scare words”: Genre as a possible bridge between L1 and L2 compositionists. *Journal of Second Language Writing*, 20(1), 24-44. <https://doi.org/10.1016/j.jslw.2010.12.001>
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18(2), 119-135. <https://doi.org/10.1016/j.jslw.2009.02.002>

- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7(1), 31-51. <https://doi.org/10.1177/02655322900700104>
- Cumming, A., Kantor, R., & Powers, D. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework*. ETS Research Memorandum.
- Doolan, St. M., & Miller, D. (2012). Generation 1.5 written error patterns: A comparative study. *Journal of Second Language Writing*, 21(1), 1-22. <http://doi.org/10.1016/j.jslw.2011.09.001>
- Eckstein, G., & Chang, R.-H. (2022). How does the language control of L1 and L2 writers develop over time in first-year composition? *Written Communication*, 39(4), 600-629. <https://doi.org/10.1177/07410883221099474>
- 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>
- Eckstein, G., Casper, R., Chan, J., & Blackwell, L. (2018). Assessment of L2 student writing: Does teacher disciplinary background matter? *Journal of Writing Research*, 10(1), 1-23. <https://doi.org/10.17239/jowr-2018.10.01.01>
- Eckstein, G., Schramm, W., Noxon, M., & Snyder, J. (2019). Reading L1 and L2 writing: An Eye-tracking study of TESOL rater behavior. *TESL-EJ*, 23(1).
- Edgington, A. (2005). "What are you thinking" understanding teacher reading and response through a protocol analysis study. *Journal of Writing Assessment*, 2(2), 125-148.
- Elder, C., Golombek, P., Weigle, S. C., Boldt, H., & Valsecchi, M. I. (2003). Effects of task and rater background on the evaluation of ESL student writing: A pilot study. *TESOL Quarterly*, 37(2), 345-354. <https://doi.org/10.2307/3588510>
- 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. [https://doi.org/10.1016/S1060-3743\(99\)80110-6](https://doi.org/10.1016/S1060-3743(99)80110-6)

- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge University Press.
- Ferris, D. R. (2009). *Teaching college writing to diverse student populations*. University of Michigan Press.
- Ferris, D. R. (2011). *Treatment of error in second language student writing*. University of Michigan Press.
- Ferris, D., Brown, J., Liu, H., & Stine, M. E. A. (2011). Responding to L2 students in college writing classes: Teacher perspectives. *TESOL Quarterly*, 45(2), 207-234. <https://doi.org/10.5054/tq.2011.247706>
- Ferris, D., Eckstein, G., & Dehond, G. (2017). Self-directed language development: A study of first-year college writers. *Research in the Teaching of English*, 51(4), 418-440.
- Ferris, D., Jensen, L., & Wald, M. (2015). Writing instructors' perceptions of international student writers: What teachers want and need to know. *CATESOL Journal*, 27(2), 55-72.
- Francis, W. N. (1954). Revolution in grammar. *Quarterly Journal of Speech*, 40(3), 299-312. <https://doi.org/10.1080/00335635409381991>
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210. [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Gillespie, P., & Lerner, N. (2000). *The Allyn and Bacon guide to peer tutoring*. Allyn and Bacon.
- Harklau, L. (1994). ESL versus mainstream classes: Contrasting L2 learning environments. *TESOL Quarterly*, 28(2), 241-272. <https://doi.org/10.2307/3587433>
- Harrington, S., Malencyzk, R., Peckham, I., Rhodes, K., & Yancy, K. B. (2001). WAP outcomes statement for first-year composition. *College English*, 63(3), 321-325. <https://doi.org/10.2307/378996>
- Hartwell, P. (1985). Grammar, grammars, and the teaching of grammar. *College English*, 47(2), 105-127. <https://doi.org/10.2307/376562>

- Huey, E. B. (1908/1968). *The psychology and pedagogy of reading*. MIT Press.
- Inhoff, A. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23, 612-624. [https://doi.org/10.1016/S0022-5371\(84\)90382-7](https://doi.org/10.1016/S0022-5371(84)90382-7)
- Institute of International Education (2017). International student enrollment trends, 1948/49-2016/17. *Open Doors Report on International Educational Exchange*. <http://www.iie.org/opendoors>.
- Jacobs, H., Zingraf, S., Wormuth, D., Hartfiel, V. F., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Newbury House.
- Janopoulos, M. (1992). University faculty tolerance of NS and NNS writing errors: A comparison. *Journal of Second Language Writing*, 1(2), 109-121. [https://doi.org/10.1016/1060-3743\(92\)90011-D](https://doi.org/10.1016/1060-3743(92)90011-D)
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Allyn & Bacon.
- Keh, C. L. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44(4), 294-304. <https://doi.org/10.1093/elt/44.4.294>
- Khalil, A. (1985). Communicative error evaluation: Native speaker's evaluation and interpretation of writer errors of oral EFL learners. *TESOL Quarterly*, 19(2), 335-351. [https://doi.org/10.1016/1060-3743\(92\)90011-D](https://doi.org/10.1016/1060-3743(92)90011-D)
- Kolln, M. (1981). Closing the books on alchemy. *College Composition and Communication*, 32(2), 139-151. <https://doi.org/10.2307/356688>
- Kolln, M., & Hancock, C. (2005). The story of English grammar in United States schools. *English Teaching: Practice and Critique*, 4(3), 11-31.
- Krashen, S. (1984). *Writing: Research, theory, and application*. Pergamon Press.
- Lance, D. M. (1977). What is "Grammar"? *English Education*, 9(1), 43-49.
- Land, R., & Whitley, C. (1989). Evaluating second-language essays in regular composition classes: Toward a pluralistic U. S. rhetoric. In D. Johnson & D. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 284-293). Longman.

- Lane, J., & Lange, E. (2012). *Writing clearly: Grammar for editing* (3rd ed.). Heinle Cengage Learning.
- Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. Routledge.
- Leonard, D. J., & Gilsdorf, J. W. (1990). Language in change: Academics' and executives' perceptions of usage errors. *Journal of Business Communication, 27*(1), 137-158. <https://doi.org/10.1177/002194369002700202>
- Lindsey, P., & Crusan, D. J. (2011). How faculty attitudes and expectations toward student nationality affect writing assessment. *Across the Disciplines: A Journal of Language, Learning, and Academic Writing, 8*. <https://doi.org/10.37514/ATD-J.2011.8.4.23>
- Lunsford, A., & Lunsford, K. (2008). "Mistakes are a fact of life": A national comparative study. *College Composition and Communication, 59*(4), 781-806.
- Matsuda, P. K. (1999). Composition studies and ESL writing: A disciplinary division of labor. *College Composition and Communication, 50*(4), 699-721.
- Matsuda, P. K. (2012). Let's face it: Language issues and the writing program administrator. *WPA: Writing Program Administration, 36*(1), 141-163.
- Matsuda, P., Saenkhum, T., & Accardi, S. (2013). Writing teachers' perceptions of the presence and needs of second language writers: An institutional case study. *Journal of Second Language Writing, 22*, 68-86. <https://doi.org/10.1016/j.jslw.2012.10.001>
- McCleary, B. (1995). Grammar making a comeback in composition teaching. *Composition Chronicle: Newsletter for Writing Teachers, 8*(6), 1-4.
- Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing, 16*(2), 82-99. <https://doi.org/10.1016/j.jslw.2007.04.002>
- Paulson, E., Alexander, J., & Armstrong, S. (2007). Peer review re-viewed: Investigating the juxtaposition of composition students' eye movements and peer-review processes. *Research in the Teaching of English, 41*(3), 304-335.

- Pelosi, A. G. (1973). What is "Grammar"? *The Modern Language Journal*, 57(7), 329-335. <https://doi.org/10.2307/324650>
- Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading*, 1, 317-339. https://doi.org/10.1207/s1532799xssr0104_2
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372-422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rifkin, B., & Roberts, F. D. (1995). Error gravity: A critical review of research and design. *Language Learning*, 45(3), 511-537. <https://doi.org/10.1111/j.1467-1770.1995.tb00450.x>
- Rubin, D. L., & Williams-James, M. (1997). The impact of writer nationality on mainstream teachers' judgments of composition quality. *Journal of Second Language Writing*, 6(2), 139-154. [https://doi.org/10.1016/S1060-3743\(97\)90031-X](https://doi.org/10.1016/S1060-3743(97)90031-X)
- Santa, T. (2006). *Dead letters: Error in composition, 1873-2004*. Hampton Press.
- Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, 22(1), 69-90. <https://doi.org/10.2307/3587062>
- Santos, T. (1992). Ideology in composition: L1 and ESL. *Journal of Second Language Writing*, 1(1), 1-15. [https://doi.org/10.1016/1060-3743\(92\)90017-j](https://doi.org/10.1016/1060-3743(92)90017-j)
- Sheorey, R. (1986). Error perception of native-speaking and non-native-speaking teachers of ESL. *English Language Teaching Journal*, 40, 306-312. <https://doi.org/10.1093/elt/40.4.306>
- Shvidko, E. (2015). Assessing preparation of mainstream composition teachers working with multilingual writers. *INTESOL Journal*, 12(2), 37-67. <https://journals.iupui.edu/index.php/intesol/article/view/20662/20347>
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: ESL research and its implications. *TESOL Quarterly*, 27, 657-677. <https://doi.org/10.2307/3587400>

- Song, B., & Caruso, I. (1996). Do English and ESL faculty differ in evaluating the essays of native English-speaking and ESL students? *Journal of Second Language Writing, 5*(2), 163-182. [https://doi.org/10.1016/S1060-3743\(96\)90023-5](https://doi.org/10.1016/S1060-3743(96)90023-5)
- Staub, A., & Rayner, K. (2007). Eye movements and online comprehension processes. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327-342). Oxford University Press.
- Sweedler-Brown, C. O. (1993). ESL essay evaluation: The influence of sentence-level and rhetorical features. *Journal of Second Language Writing, 2*(1), 3-17. [https://doi.org/10.1016/1060-3743\(93\)90003-L](https://doi.org/10.1016/1060-3743(93)90003-L)
- Tomiyana, M. (1980). Grammatical errors communication breakdown. *TESOL Quarterly, 14*(1), 71-79. <https://doi.org/10.2307/3586810>
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning, 46*(2), 327-369. <https://doi.org/10.1111/j.1467-1770.1996.tb01238.x>
- Truscott, J. (1999). The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing, 8*(2), 111-122. [https://doi.org/10.1016/S1060-3743\(99\)80124-6](https://doi.org/10.1016/S1060-3743(99)80124-6)
- Weigle, S. C., Boldt, H., & Valsecchi, M. I. (2003). Effects of task and rater background on the evaluation of ESL student writing: A pilot study. *TESOL Quarterly, 37*(2), 345-354.
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*(3), 475-482. <https://doi.org/10.1007/BF03395630>
- Williams, J. (1981). *Style: Ten lessons in clarity and grace*. Scott Foresman.
- Wolfe, E. (2005). Uncovering rater's cognitive processing and focus using think-aloud protocols. *Journal of Writing Assessment, 2*(1), 37-56.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19*, 79-102. <https://doi.org/10.2307/3586773>

Appendix A: Sample Stimuli Paragraph

The biggest problem I face in writing is organizing my thoughts. Perhaps due to my strong disliking of reading book (S/P) as a child, I struggle tremendously with organizing my thoughts and displaying them into the (determiner) words; I face this same problem even in verbal communication. Prior to starting an essay (missing comma) I fill my head with so many different idea (S/P) that I consider brilliant, and start to design a (determiner) illustration of it (vague pronoun) into words and paragraphs. Unfortunately, often times this turns out to be nothing but my literary fantasy. When I am actually writing an essay they (vague pronoun) become a mash-up of vague ideas that don't even seem to relate on (preposition) one another. They are just a random collection of ideas. By the end of it (vague pronoun), the reader might not even remember what the thesis of the essay was and clueless as to what message I was trying to convey.

Appendix B: Rubric used for Paragraph Assessment

Content	Organization	Vocabulary	Grammar
7 • Thorough development of thesis • Substantive and detailed	7 • Excellent paragraph organization • Clear thesis statement or main idea	7 • Very sophisticated vocabulary • Excellent choice of words with no errors	7 • No spelling errors • No punctuation errors • No major errors in word order or complex structures
6 • No irrelevant information • Interesting	6 • Excellent use of transition words	6 • Excellent range of vocabulary • Academic register	6 • Excellent sentence variety • No errors that interfere with comprehension
5 • Good development of thesis • Fairly substantive and detailed	5 • Good paragraph organization • Clear thesis statement or main idea	5 • Somewhat sophisticated vocabulary • Good choice of words with some errors that don't obscure meaning	5 • Minor spelling errors in less frequent words • No more than a few punctuation errors
4 • Somewhat interesting	4 • Good use of transition words	4 • Adequate range of vocabulary • Approaching academic register	4 • Occasional errors in awkward order or complex structures • Good sentence variety • Almost no errors that interfere with comprehension
3 • Some development of thesis • Not much substance or detail • Some irrelevant information	3 • Some coherent organization • Minimal thesis statement or main idea	3 • Unsophisticated vocabulary • Limited word choice with some errors obscuring meaning	3 • Some spelling errors with less frequent words • Several punctuation errors • Errors in word order or complex sentences
2 • Somewhat uninteresting	2 • Occasional use of transition words	2 • Repetitive choice of words • No resemblance to academic register	2 • Little sentence variety • Some errors that interfere with comprehension
1 • No development of thesis • No substance or details • Substantial amount of irrelevant information	1 • No coherent organization • No thesis statement or main idea	1 • Very simple vocabulary • severe errors in word choice that often obscure meaning • no variety in word choice	1 • Spelling errors even in frequent words • Many punctuation errors • Serious errors in word order or complex structures
0 • Completely uninteresting	0 • No use of transition words	0 • no resemblance to academic register	0 • no sentence variety • frequent errors that interfere with comprehension