

An Exploratory Study of Grammarly in the Language Learning Context: An Analysis of Test-Based, Textbook-Based and Facebook Corpora

Daniel Bailey

Konkuk University Glocal Campus, South Korea

Andrea Rakushin Lee*

Konkuk University Glocal Campus, South Korea

Abstract

Different genres of writing entail various levels of syntactic and lexical complexity, and how this complexity influences the results of Automatic Writing Evaluation (AWE) programs like Grammarly in second language (L2) writing is unknown. This study explored the use of Grammarly in the L2 writing context by comparing error frequency, error types and writing complexity for university admission test essays, textbook-based descriptive essays, social network site (SNS) posts, and SNS comments. Several findings were revealed. Punctuation, grammar, vocabulary, and spelling mistakes were the most common error types across all corpora. In particular, determiner errors such as article usage were most frequent while differences existed across corpora related to incomplete sentences, run-on sentences, and spelling. Test-based and textbook-based compositions consisted of longer sentences with less lexical variation than SNS-based writing. SNS posts and comments produced greater noun, verb, and modifier variation. After applying Grammarly corrections, SNS posts resulted in the greatest clarity and this was attributed to shorter sentence length and simpler word-choice than textbook-based and test-based writing. Grammarly was more appropriate for local surface-level errors (e.g. articles, preposition, and verb-noun agreement) while instructors are needed for issues related to awkward wording and cohesion. Pedagogical implications in the English as a Foreign Language (EFL) context are given.

Keywords: corrective feedback, automatic writing evaluation, Grammarly, second language writing, syntactic complexity, and lexical density, Facebook

* Email: andreelee1216@gmail.com.

Introduction

While controversy currently exists concerning the application of Automatic Writing Evaluation (AWE) in test settings, there is little argument to the time-saving benefits afforded by computer generated feedback. The accuracy of AWE programs increases each year, and there is a growing debate among language instructors on whether or not to allow students to use these tools to improve their L2 writing. Research has found that AWE programs are adequate at detecting errors (Burston, 2008) at rates approaching the effectiveness of teacher corrections (Nadasdi & Sinclair, 2007). AWE programs can increase linguistic writing quality and engagement (Gauthier, 2013). This observed improvement in accuracy is accompanied by heightened levels of motivation and confidence (Potter & Fuller, 2008). Students are satisfied with corrective feedback from programs like Grammarly, especially in conjunction with a human rater (O'Neill & Russel, 2019), and while research shows the utility of programs like Grammarly, exactly how AWE programs perform across different L2 writing genres or error types is still unclear.

Automatic writing evaluation has undergone scrutiny over the years. Critics cite shortcomings with AWE concerning the poor validity of scoring tests (Chung & Baker, 2003) and unreliable assessment of social communication (Ericsson, 2006). Further condemnation persists because AWE programs have difficulty recognizing deeper level, global errors, related to such things as cohesive links and errors in factual content (McGee, 2006). Moreover, overreliance on AWE software prevents students from engaging language learning strategies such as looking up unknown words or asking for help. According to the comprehensible output hypotheses, what learners do when pushed is responsible for some amount of language acquisition (Swain, 2000). Language learning occurs in part by clarifying meaning during the writing process, but this clarification entails less cognitive engagement with AWE. Consequently, students spend less time looking up unknown words or discussing corrective feedback with more knowledgeable peers.

Despite the opposition, there are undeniable advantages for L2 writers and instructors when implementing AWE. The accuracy of AWE programs continues to improve in parallel with decreasing costs, and time saved by teachers can be allocated to other student needs. Furthermore, student-centered feedback with AWE addresses some of the challenges common in differentiated EFL/ESL instruction. Scaffolding feedback, according to language proficiency, allows the L2 writer to negotiate meaning within their zone of proximal development. Krashen's input hypothesis (1985) posits that language learners progress in their knowledge of the language when they comprehend language input slightly above their current level. AWE feedback provides insight into the learner's next stage of language acquisition because the revisions address what the user composed on their own. Students receive input from AWE at an $i+1$ level, where i is learner's interlanguage and $+1$ is the next stage of language acquisition. The benefits of AWE depend on how writing is defined and the content that is written (Ware, 2011).

Understanding the efficacy of AWE platforms with different writing conditions helps justify its use in the EFL context. Factors such as L2 writing anxiety, available time to write, academic stakes,

and the audience are considered here to be conceptually unique for test-based, textbook-based, and SNS-based writing genres. Therefore, text from each of these genres is analyzed.

The test-based corpus in the current study came from university entrance exam essay questions. Academic stakes are higher with test-based writing, and there is less time for pre-writing, when-writing, and post-writing strategies. Time limits with test-based writing prevent lengthy pre-writing planning strategies and post writing reflective strategies. High stakes conditions with writing tests also influence L2 writing anxiety.

Textbook-based writing produced in the classroom using time-consuming writing strategies results in linguistic differences when compared to other forms of composition. Textbook-based writing provides more time for more L2 writing strategies (e.g., mind-mapping, brainstorming, outlining, and reflection) and allows students to use models of correct composition. However, lower stakes with textbook-based writing may mean less motivation to perform well on writing tasks compared to university admissions test conditions.

Social network sites like Facebook allow students to practice real-world writing when communicating. The terms Facebook and SNS are used interchangeably throughout this manuscript and refer to informal real word written communication for social purposes. Students share their composition with SNS writing, and this elicits motivation by comparison but also anxiety from fear of looking foolish in front of their peers. SNS writing on platforms like Facebook can be categorized as main posts and comments/replies (*hereafter* comments). Main posts are similar to diary or journal entries while comments often entail shorter statements. Bailey et al. (2017) documented the patterns in participation among students in a Facebook for language learning program and recognized differences in sentence complexity between main posts and replies. The different parameters with each of these writing tasks are expected to produce different error types, error rates, and writing complexity.

Grammarly, an AWE platform that identifies over 300 types of writing errors, was used in the current study to compare error types and writing complexity for different EFL writing genres. No argument is being made that Grammarly is better than other AWE platforms. This study provides insight into the usefulness of AWE programs like Grammarly as an L2 writing tool by recognizing which types of errors Grammarly correctly identifies, and how those identified errors vary across different EFL writing tasks.

Literature Review

AWE can assist both low- and high-performing students with their accuracy and save instructors valuable time. The current study posits that AWE programs like Grammarly are valuable L2 writing tools that can help educators grade writing and help students revise compositions. The following literature review explores the current state of AWE research and the role AWE has in the context of L2 writing.

Automatic Writing Evaluation

The development of AWE software has enabled L2 learners to receive feedback on language and content in addition to automated scores. Benefits to writing accuracy were recognized in AWE studies (Anson, 2006; Dikli, 2010). These results drive the popularity of integrating automatic corrective feedback in the classroom (Li, Link, & Hegelheimer, 2015; Zhang & Hyland, 2018). Researchers continue to find AWE helpful for measuring linguistic accuracy (Li et al., 2015). Li et al. (2015) used AWE error reports to explore the role of Criterion in an ESL writing curriculum. By looking at error types, they found that automatic writing evaluation software led to increased revisions and improved accuracy. Li et al. (2015) required students to meet a benchmark score through AWE, which resulted in higher participation with writing practice and a heightened motivation to write.

AWE has the potential to liberate teachers by freeing up time; however, the capacity of AWE to provide useful feedback is still under debate (Anson, 2006). Stevenson and Phakiti (2014) set out in a meta-analysis of AWE research carried out in writing classrooms. Results indicated only modest evidence that AWE feedback had a positive effect on the quality of the texts students produced and little evidence about whether AWE was associated with more general improvements in writing proficiency. They suggest AWE feedback is more beneficial when combined with teacher feedback, and make a call for future research to examine AWE effectiveness in ESL and EFL settings. Li et al. (2015) make a similar call for research by noting a lack of evidence for the use of AWE corrective feedback in ESL writing classrooms.

Most AWE research has focused on a written product with less attention to the revision process (Zhang & Hyland, 2018). To address this gap in research, El Ebyary and Windeatt (2010) investigated whether the use of computer-based feedback using the AWE program Criterion would influence attitudes towards feedback, the writing process, and writing product. Through pre/post-survey analysis, interviews, and holistic scores from Criterion, they found that students used few pre-planning strategies and perceptions to feedback improved after using Criterion. The researchers recognized that teacher intervention is critical for motivation and program success and recommend the integration of AWE into conventional writing instruction. Drawbacks to AWE in their program included cheating and stagnant levels of revisions.

Zhang and Hyland (2018) conducted a study of two Chinese students by comparing AWE (with Criterion) and teacher feedback over the course of a semester. They found that AWE provided by Criterion tended to highlight errors instead of correcting them. Explicit corrective feedback has shown to be superior to simply bringing awareness to errors through indirect feedback (Chandler, 2013; Katayama, 2007), but simple awareness of the location of mistakes still leads to better writing outcomes (Ferris, 2004). AWE with Criterion provided a substantial amount of marginal comments, which could be considered a praising attribute because identifying errors is an early step to overcoming similar mistakes in future writing. However, the two Chinese students in Zhang and Hyland's (2018) study felt the automatic feedback was overwhelming and repetitive.

A current debate exists around AWE regarding effectiveness for students at different L2

proficiency levels. A certain amount of minimum proficiency is necessary to negotiate feedback delivered in the L2, but where that level rests is unknown. AWE systems may benefit more competent L2 writers more than less competent ones because stronger writers have the communicative and rhetorical understanding of the target language allowing them to make more appropriate use of the AWE feedback. In line with this argument, Dikli (2010) suggests some AWE is too overwhelming for low-performing writers because such writers may require basic metacognitive knowledge (Liao, 2016) or possess a base knowledge of the L2 (Caveleri & Dianati, 2016; Hoang & Kunnan, 2016). Other researchers have found AWE benefited low-performing writers (Chen & Cheng, 2008; O'Neill & Russell, 2019) by giving them input on correct writing form.

Some criticisms of AWE relates to a lack of human interaction. Dikli (2010) recommends adding mechanisms to AWE that promote individualized feedback, such as providing formative and cumulative feedback and increasing the number of praise messages. Dikli (2010) explored the nature of feedback that ESL students received on their writing either from a human rater or AWE. Teacher feedback was shorter and more to the point while AWE feedback was redundant (e.g., repetitive information) and not student centered. Feedback from the teacher was late or non-existent. According to Dikli (2010), it is necessary to consider effective methods to analyze large number of essays and provide individual feedback. The current study posits that AWE should be used to address local error types, such as surface-level revisions (Stevenson, 2016), while teacher feedback can focus on global issues.

Stevenson (2016) investigated the ways AWE is used or could be used as an instructional tool in the writing classroom and provides an overview of what is known about AWE's application as a class-instruction tool. Specifically, Stevenson's (2016) meta-analysis looked at Criterion and My Access! studies and identified three key constructs in the relevant literature relating to the integration of AWE in the writing classroom which were purpose, action, and use. The purpose of AWE in the classroom was primarily to save teachers time (Chen & Chang, 2008; Li et al., 2015), promote learner autonomy (Chen & Chang, 2008; Grimes & Warschauer, 2010), develop writing processes (El Ebyary & Windeatt, 2010), and raise awareness (Li et al., 2015). Integration of AWE in the classroom can be made possible through scaffolding (Grimes & Warschauer, 2010), embedding in classroom instruction (Li et al., 2015), and assessment and exam preparation (Chen & Chang, 2008).

Caveleri and Dianati (2016) investigate student perceptions of Grammarly in terms of the Technology Acceptance Model (TAM). Overall, students reported that the explanations Grammarly provided helped them understand grammar rules. Grammar books and paper-based exercises on photocopied handouts are portable but lack the direct interactivity with students, which online grammar checkers can provide. Furthermore, feedback from Grammarly "led to reflection about grammar that may not have occurred otherwise" (Caveleri, p. 233). Through a mixed method exploratory design that compared responses from students receiving feedback from Grammarly and students receiving feedback from the teacher, O'Neil and Russell (2019) found that students using Grammarly

responded more positively and enjoyed AWE significantly more than instructor feedback. Students in the Grammarly group were satisfied with the amount of time that was spent on feedback versus the teacher-feedback group. Furthermore, students in the Grammarly group were more likely to state that they had received useful feedback. The researchers recommend that “the program [Grammarly] is used in conjunction with academic learning advisor input as the program is currently not accurate enough for independent use to be justified” (O’Neil & Russel, p. 42). It should be noted that both feedback groups were satisfied with the feedback received but the Grammarly group was significantly happier. O’Neil and Russell (2019) recognized a weakness with AWE related to the inaccuracy of some feedback, and they feel further investigation is needed to identify the errors which Grammarly most frequently missed or misidentified.

The impact of AWE integration on key stakeholders requires a clearer understanding of the context in which the AWE system is used and the content of what is written. To help understand this integration, our study explored the extent that AWE effectively assessed error type, error frequency, and sentence complexity in the EFL context and how that assessment differed across conceptually unique L2 writing genres. This study first identified similarities and differences in error types across L2 writing genres. Next, syntactic complexity and lexical variation were measured. The research questions of the current study were formulated as follows:

1. What are the similarities and differences in error-types among admission test-based writing, descriptive textbook-based essay writing, SNS posts, and SNS comments?
2. What are the similarities and differences in syntactic complexity and lexical variation among admissions test-based writing, textbook-based essay writing, SNS posts, and SNS comments?

Methods

This corpus analysis study compared error types, error frequency, lexical variation and syntactic variation across four genres of writing - university admission writing tests, textbook-based essays, social media posts, and social media comments. Table 1 displays information on the four sets of corpora.

Table 1
Corpora Word Count

Corpus	<i>M</i>	<i>SD</i>	Words	N
Textbook	2,375	554.03	52,250	22
Test	369	43.87	36,940	100
Posts	711	326.05	29,151	41
Comments	736	431.04	19,136	26

Corpora

Writing for the test-based corpus came from Yonsei English Learners' Corpus (YELC). Undergraduate students were administered the 2011 Yonsei English Placement Test (YEPT). The YEPT writing section lasted 60 minutes and students completed two questions. Question One called for free writing and Question Two asked students to write an argumentative essay. The YELC classifies composition quality according to the Common European Framework of Reference (CEFR) criteria. All students in the test-based group were between the age of 18 and 19 and were equally parsed according to gender (M = 50, F = 50) and further parsed according to L2 proficiency. Only B1 (N=35), B1+ (N=40), and B2 (N=35) essays were analyzed in order to represent similar proficiency levels to the students who produced the textbook-based essays and SNS-based writing. The following is an example passage taken from a YELC composition:

To be honest, I think physical punishment should be allowed in schools. In fact besides physical punishment there are other kinds of punishment such as additional homework. But these other types of punishment have not been effective in school...

Students in the textbook-based writing corpus group followed process writing instruction to complete a series of writing tasks from the book *Writing from Within 2* (Kelly & Gargagliano, 2011). The corpus of textbook-based writing consisted of 52,250 words from 22 English majors attending an L2 writing class. The 22 (M = 4, F = 18) English majors in the textbook writing group were on average 22.1 years old (SD = 1.48). Writing activities involved traditional writing instruction in which the instructor introduced a writing topic, used modeling and writing templates to help students develop ideas, helped students begin first drafts, and discussed methods to improve writing. The six descriptive writing tasks administered had similar difficulty with one another and provided equal amounts of scaffolding via modeling and brainstorming activities. Writing collected came from the first drafts. Task one asked students to describe their favorite place; task two asked students to describe their personal qualities; task three asked students to describe an invention; task four asked students to describe something that changed their life; task five asked students to describe their dream job; task six asked students to describe their personal goals. First drafts from these assignments were compiled for each student and then processed through Grammarly. The following is an example passage from a composition written by a student in the textbook corpus group:

Minji has a curious personality. She is always interested in working in many different fields. Regarding her most likely personality, a professor would be a great future job for her. She likes to teach, research, and give counsel. First of all, Minji likes teaching. She feels worth when...

SNS-based compositions were parsed into main posts and comments. At least 300 words were

set as the minimum number of words a student had to produce for inclusion into either the main post or comment corpus. The 41 (M = 18, F = 28) English majors in the Facebook group were on average 22.4 years old (SD = 1.36). Of the 41 students who contributed main posts, only 26 contributed at least 300 words in their comments (Male = 10, Female = 16). Students in the Facebook group were attending a Multimedia English course. According to the International English Language Testing System (IELTS) rubric, students in the textbook and SNS groups had similar L2 proficiency ranging from B1 to B2 CEFR Levels. Examples of Facebook posts and comments are displayed in Table 2. Proficiency levels were assessed by the researcher-instructor embedded with participants during a 16-week semester.

Table 2
Examples of Actual Main Posts, Comments and Replies

Main Post

Hello everyone? How was your weekend?

I had a good time. I met my friends at the Igseon-dong station. There were so many people. There stores are great. It's very unique. And, it is a single-story building. The interior was good. At first, we tried to eat dumplings. But...(Continued)

Comment 1

(reflective) *Wow ~ It was so delicious! I will go there!*

Reply

Yes, it will be good!

Comment 2 *Oh! I went to Igseon-dong last week!! The alley was so beautiful. I saw the restaurant! But I didn't go. There were many people waiting! I went to another restaurant! I ate spaghetti!!*
(substantive) *There was no restaurant sign. I want to go to this place again.*

Procedures

First, the text from each corpus needed prepared for analysis. Methods for preparing documents differed among corpora because of varying word count among group members. For the textbook and SNS groups, each students' writing was analyzed separately from other students in their corpus group. In the textbook group, all of the writing tasks from the same student were combined and then analyzed so there was a total of 22 Grammarly reports produced. Similar steps were taken for the SNS post and SNS comment group creating 41 and 26 Grammarly reports respectively. For the test-based writing group, students were aggregated into groups of five and Grammarly reports were created for each of these groups, which resulted in a total of 20 Grammarly reports.

Grammarly performance reports provided analytics on the error-types, syntactic complexity, and overall writing scores. Grammarly offers free automatic feedback on critical errors while the commercial version provides feedback on advanced errors and allows users to download analytics (i.e., performance reports). The commercial version was used for this study. Figures 1 and 2 display samples of a performance report.

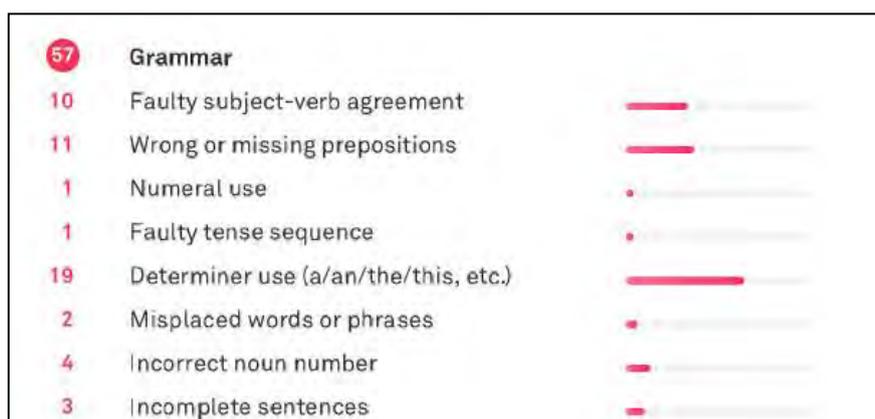
Figure 1

Example of Grammarly Performance Report Part 1



Figure 2

Example of Grammarly Performance Report Part 2

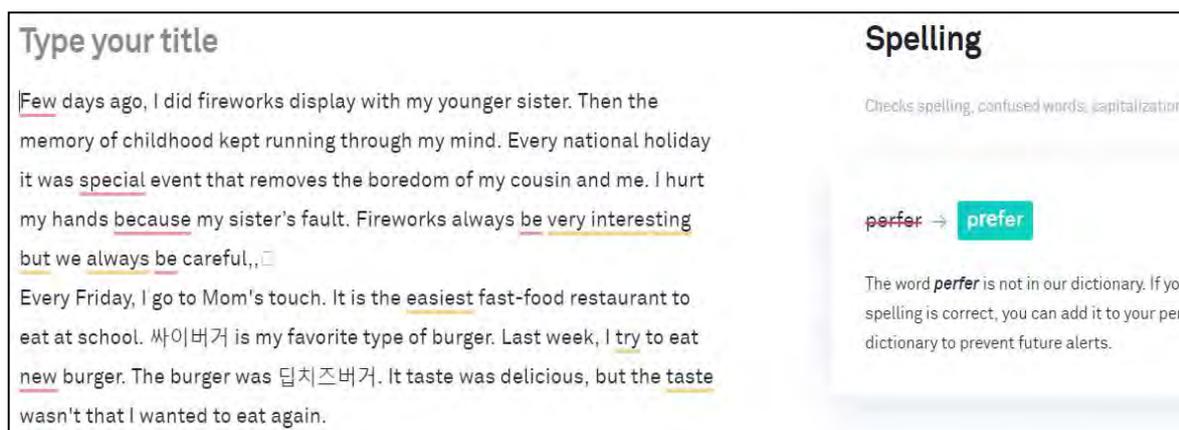


When using Grammarly, the composition is displayed on the left side of the screen and corrections can be made in real time. Figure 3 shows an example of the text editor panel for

Grammarly. This information is provided to the user in real time so that corrections can be applied while composing text.

Figure 3

Example of the Grammarly Text Editor



While Grammarly identifies over 300 error types, this study only analyzed error-types that occurred 0.3 times out of a hundred words for at least one of the three writing sets. Table 3 displays error categories analyzed by Grammarly along with brief descriptions and examples of some error-types. Errors that occurred less frequently were aggregated into error groups (e.g., punctuation, grammar, conventions).

Table 3
Error-Type Categories

Error Type	Description	Feedback Given by Grammarly
Punctuation	Corrects missing and misused punctuation	It appears that you are missing a comma before the coordinating conjunction <i>but</i> in a compound sentence.
Complex or Compound Sentence	<i>e.g., Run on sentences</i>	
Comma in Clause	<i>Well [,] I am not sure about that</i>	
Grammar		
Determiner	<i>I think [the] bottle is so cute.</i>	
Incomplete Sentence	<i>Never on time.</i>	
Subject-Verb	<i>I drove [rode] my bike to school.</i>	
Preposition Error	<i>He goes on his work.</i>	
Variety	Provides synonyms for repeated words	The word <i>order</i> appears repeatedly in this text. Consider using a synonym in its place.
Convention	Checks spacing, capitalization, and dialect-specific spelling	The numeral 2 is used instead of the word spelled out. Consider spelling out the number.
Spelling	Corrects misspellings, misused words, and capitalization	
Conciseness	Eliminates wordiness and redundancy	It appears that <i>Actually</i> , may be unnecessary in this sentence.

In addition to error-types, certain properties of syntactic and lexical diversity were analyzed.

Grammarly produced syntactic analytics for word and sentence length. Furthermore, unique words and rare words were also identified and compared among writing sets. Lu's (2012) lexical complexity analyzer was used to identify noun, verb, and modifier variation. Lexical sophistication measures the proportion of relatively unusual or advanced words in the learner's text (Read, 2000), and lexical variation (i.e., lexical diversity) refers to the range of a learner's vocabulary as displayed in their language use. Table 4 displays the indices and the formulas used to calculate values.

Table 4
Lexical and Syntactic Indices

Word Length	Total words / Total Characters
Sentence Length	Total Sentences / Total Words
Noun Variation*	Different Nouns / Total Nouns
Verb Variation*	Different Verbs / Total Verbs
Modifier Variation*	Different Modifiers (adjective + adverbs) / Total Modifiers
Unique Words	Measures vocabulary diversity by calculating the percentage of words used only once in your document
Rare Words	Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

Note: *Variation indices were computed using Lu (2013) lexical complexity analyzer.

Data Analysis

The software package SPSS 24.0 was used to analyze data. Analysis of variance (ANOVA) is a common statistical procedure used to determine whether any statistically significant difference between the means of two or more unrelated groups. For research question one, mean score comparison with one-way ANOVA was carried out to identify statistically significant differences in error types among test-based, textbook-based, and SNS-based corpora. For research question two, a similar mean score comparison with one-way ANOVA was used to identify statistically significant differences for lexical and syntactic complexity among the writing sets.

Results and Discussion

Answering research question one begins by describing the mean scores for error types recognized by Grammarly (Table 5). Details are given for the most frequent errors and statistically significant differences in error types between genres. Research question two goes on to describe the lexical and syntactic similarities and differences across genres.

Table 5
Descriptive Statistics for Error Categories

Error Category		<i>M</i>	<i>SD</i>	Error Category		<i>M</i>	<i>SD</i>
Errors Per 100 Words	1	8.80	3.24	Variety^a	1	0.63	0.27
	2	13.13	2.37		2	1.24	0.44
	3	10.85	4.01		3	0.69	0.51
	4	11.83	3.14		4	0.48	0.42
	Total	11.05	3.56		Total	0.73	0.50
Grammar	1	3.12	1.40	Conventions	1	1.01	1.03
	2	4.81	1.17		2	0.72	0.78
	3	2.71	1.34		3	0.52	0.68
	4	3.28	1.31		4	1.09	1.79
	Total	3.32	1.50		Total	0.79	1.14
Determiner	1	1.36	0.65	punctuation	1	1.02	0.46
	2	2.68	0.61		2	0.99	0.22
	3	1.34	0.83		3	1.13	0.80
	4	1.45	0.75		4	2.29	1.22
	Total	1.62	0.89		Total	1.36	0.95
Incomplete Sentences	1	0.25	0.19	Comma in Clause	1	0.34	0.22
	2	0.24	0.12		2	0.42	0.17
	3	0.37	0.35		3	0.30	0.41
	4	0.83	0.67		4	1.04	0.90
	Total	0.43	0.46		Total	0.51	0.59
Faulty Subject-Verb Agreement	1	0.32	0.39	Punctuation in Compound/Complex Sentence	1	0.51	0.29
	2	0.57	0.32		2	0.48	0.19
	3	0.17	0.20		3	0.59	0.42
	4	0.18	0.24		4	0.65	0.53
	Total	0.27	0.32		Total	0.57	0.39
Preposition	1	0.45	0.19	Spelling	1	1.46	0.88
	2	0.46	0.15		2	3.06	1.04
	3	0.26	0.22		3	1.96	1.76
	4	0.30	0.31		4	2.34	1.85
	Total	0.35	0.24		Total	2.15	1.60
Conciseness (Wordiness)	1	0.33	0.18				
	2	0.24	0.19				
	3	0.54	0.46				
	4	1.16	0.74				
	Total	0.59	0.58				

Note: Textbook-based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4; Variety^a refers to repeated word choice errors.

Grammar Errors

Results from answering research question one showed students committed grammar mistakes the most out of the Grammarly error-type groups. Of the four corpora, students in the test-taking condition produced the most grammar mistakes followed by Facebook posts and comments and then text-book based writing. As indicated in Table 6, these differences were statistically significant. Students committed more mistakes under time-constrained testing conditions.

Among error types in the grammar category, determiner errors were the most common for all groups, with the test-based writing conditions resulting in twice as many determiner errors than the other groups. Determiners indicate whether or not a noun is definite or indefinite and are common

among South Korea L2 writers (Lee, Chodorow, & Gentile, 2016). The most common determiner errors related to articles, which are difficult for L2 English learners (Han, Martin, & Leacock, 2006). One of the most difficult challenges for L2 writers is mastering the use of English articles (Han et al., 2006). Correctly using definite and indefinite articles can be difficult in different linguistic environments and even with the inclusion of instructor-provided corrective feedback (Ferris, Chaney, Komura, Roberts, & McKee, 2000). Grammarly drastically reduces article and other determiner errors.

There were a number of other grammar error-types that differed in ratio among the writing sets. Facebook comments had three times as many incomplete sentence errors than either the test or textbook-based writing groups and twice as many as the Facebook main posts. Incomplete sentences are common in SNS communication because ideas can be *completed* with non-textual multimodal (e.g., photos, links, or video) communication (Bailey et al., 2017). Students write shorter sentences in Facebook comments, and these sentences can often be incomplete statements referencing a point of discussion located further up the discussion chain, or in comments that are expressing an emotion, and therefore more incomplete sentence error and incomplete comparison errors occur.

Table 6
Bonferroni Post Hoc Analysis of Error Categories across Genre

Error Type			<i>MD</i>	<i>SE</i>	<i>p</i>
Grammar	2	1	1.69	0.41	.000
		3	2.10	0.36	.000
		4	1.53	0.39	.001
Determiner	2	1	1.32	0.23	.000
		3	1.33	0.20	.000
		4	1.23	0.22	.000
Incomplete Sentence	1	4	-0.58	0.12	.000
	2	4	-0.59	0.12	.000
	3	4	-0.46	0.10	.000
Incorrect Noun	2	3	0.11	0.04	.017
Faulty Subject-Verb Agreement	2	1	0.24	0.09	.038
		3	0.40	0.08	.000
		4	0.39	0.08	.000
Preposition	3	1	-0.18	0.06	.017
		2	-0.20	0.06	.011
Conciseness	4	1	0.83	0.14	.000
		2	0.92	0.14	.000
		3	0.62	0.12	.000
Variety	2	1	0.61	0.14	.000
		3	0.54	0.12	.000
		4	0.76	0.13	.000
Spelling	1	2	-1.59	0.47	.006
		4	-0.88	0.44	.304
Punctuation	4	1	1.27	0.23	.000
		2	1.29	0.24	.000
		3	1.15	0.20	.000
Comma in Clause	4	1	0.70	0.15	.000
		2	0.62	0.15	.001

Note: Textbook-based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4

Incorrect noun choice was the next error-type in the grammar category that revealed a significant difference across groups. Test-based writing resulted in more noun choice mistakes than Facebook posts. Test-based writing consisted of longer and rarer words than Facebook writing. A similar pattern of more errors occurring under test-based conditions appears with subject-verb agreement (e.g., He *have* to go – He *has* to go) and prepositions (e.g., *at* Monday – *on* Monday). The students writing essays committed more grammar mistakes whether under the time time-dependent test-based condition or time-independent textbook-based condition, and this was attributed to the inclusion of more rare words and longer sentences.

AWE software like Grammarly can be used to address what Ferris (1999) termed *treatable* errors such as verb tense and form, subject-verb agreement, article usage, plural and possessive noun endings, and sentence fragments, while writing instructors can provide an informed complementary feedback approach for identifying errors within a structural or textual whole (Hamilton, 2015).

Incomplete reconstruals are common errors committed by East Asian writers in countries like

China (Liardet, 2015). Grammarly was able to consistently identify incomplete reconstructions like *the understand between people* and correct them with feedback like *the understanding between people*.

Conciseness, Convention, and Spelling Errors

Error types concerning conciseness, or wordiness, also revealed differences in error ratios across groups. Facebook posts resulted in twice as many conciseness errors and Facebook comments resulted in four times as many than textbook-based and test-based writing. The reasoning for this was attributed to casual language which is common in Facebook posts. Weak adjectives such as “like”, “really”, and “very” were frequently found in Facebook writing. To compensate for the overuse of *weak* words, Grammarly provided more sophisticated alternatives.

Test-based and Facebook posts revealed more writing in the passive voice. While not incorrect, the passive voice is often considered a bad writing habit. Active voice is preferred because it makes writing stronger and more direct. Writing from test-taking conditions resulted in twice as many variety, or word choice, errors. For example, if the student wrote about smoking in public then the Grammarly program would offer alternatives to “smoking” due to the overuse of the word. Conciseness, passive voice, and word-choice errors are not technically mistakes, but characteristics of writing quality that Grammarly attempts to improve.

Convention errors (e.g., capitalization and spacing errors) occurred in all sets of writing. Facebook posts produced the least amount of convention errors but not at a statistically significant level. With Facebook, the higher rate of accuracy may be attributed to asynchronous writing and access to devices that correct automatically. Spelling errors (spelling and capitalization of proper nouns) were the second most frequently occurring error type category recognized by Grammarly. Textbook-based writing resulted in the least amount of spelling mistakes, while Facebook posts and test-based writing produced the most. Many of the words misspelled in the test-based group were high-frequency words (i.e., most common 1000 words) such as *thought* (thought), *frend* (friend), and *raining* (raining) and were attributed to the time-dependent nature of the task. Furthermore, students in the test-based condition had no access to spell-checkers or dictionaries which likely contributed to these convention and spelling errors. In addition to misspelled words, Facebook posts revealed numerous examples of proper nouns starting with lowercase letters.

The four writing sets exhibited high levels of spelling errors which is also common among L2 writers (Lee et al., 2016). SNS writing had more spelling mistakes related to unknown words, which is expected with greater topic variety and inclusion of Korean names (*bulgogi*, meaning meat) and locations (*Mapo Naru*, a popular restaurant in Seoul). Instructors are recommended to provide examples of when to avoid recommendations from Grammarly when country-specific terms such as unknown proper nouns like names and locations are used by the writer.

Punctuation Errors

Punctuation errors were the third most common error category across the four corpora. Facebook comments resulted in twice as many punctuation errors as the other writing sets and this was attributed to mistakes with comma usage in clauses. Facebook comments use ellipsis to signify pause and emphasize an emotion not adequately captured by an exclamation point. For instance, in comments, students would write, “Oh my... how cute!” Here, we see an utterance expressing engagement “Oh my...” followed by excitement as indicated by “how cute!” These are examples of a SNS genre-specific writing styles and not necessarily writing errors.

Test-based and textbook-based writing was more complex than spoken language (Halliday, 1989), and this is reflected in the genre of SNS writing. Facebook writing was grammatically less complex than test-based and textbook-based writing. Incomplete sentence errors were more frequent in Facebook. The use of Facebook benefits language learners and creates communities to practice language learning in informal environments (Newgarden, 2009) which results in the use of casual languages in the form of utterances (Bailey et al., 2017). Frequent use of utterances in writing resulted in shorter sentences and greater frequency of punctuation errors related to incomplete sentences.

Writing that occurred on Facebook had more punctuation mistakes. These included punctuation mistakes with sentence-endings and punctuation mistakes within clauses. Communication with SNS is similar to the utterances that occur during face to face communication than writing that occurs through textbook-based essay writing and therefore Facebook writing was more casual than writing emanating from process writing instruction. This informal nature with Facebook writing allows for more freedom with punctuation use. Examples of loose punctuation rules applying to Facebook writing include triple ellipses [...] inside sentences to reference pauses, and double exclamation points at the end of sentences to emphasize emotion.

Research question two investigated the similarities and differences in lexical and syntactic diversity across the four sets of writing. Table 7 displays mean scores and Table 8 shows results from Bonferroni posthoc analysis of the measured indices. Statistically significant differences were recognized in each of the measured syntactic and lexical indices. Students in the textbook-based group wrote longer sentences and words while students in the SNS-based group used a wider variety of nouns, verbs, and modifiers (e.g., adjectives and adverbs).

Table 7
Lexical and Syntactic Indices

Indices	Corpus	M	SD	Indices	Corpus	M	SD
Different Words/Random 50 Words	1	37.2	1.32	Adverb Variation	1	.05	.007
	2	37.1	1.32		2	.05	.007
	3	37.3	1.67		3	.07	.022
	4	37.1	1.78		4	.08	.041
Verb Variation	1	.10	.01	Modifier Variation	1	.13	.015
	2	.10	.01		2	.13	.011
	3	.14	.04		3	.17	.034
	4	.16	.07		4	.19	.073
Noun Variation	1	.46	.05	Word Length	1	4.3	0.09
	2	.42	.04		2	4.6	0.25
	3	.61	.08		3	4.1	0.19
	4	.69	.15		4	3.8	0.79
Adjective Variation	1	.08	.01	Sentence Length	1	10.8	1.32
	2	.08	.01		2	11.9	1.17
	3	.10	.02		3	9.3	1.91
	4	.11	.04		4	6.8	1.02

Note: Textbook-Based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4

Table 8
Bonferroni Post Hoc Analysis of Lexical and Syntactic Complexity Indices

			MD	SE	p				MD	SE	p
Verb Variation	1	3	.037	.010	.003*	Modifier Variation	1	3	.041	.011	.002
			.057	.011	.000**				4	.062	.012
	2	3	.036	.011	.006*		2	3	.041	.012	.004*
			.056	.012	.000**		4	.062	.013	.000**	
Noun Variation	1	3	.153	.025	.000**	Word Length	1	4	0.538	0.121	.000**
			.237	.027	.000**				2	3	0.497
	2	3	.194	.026	.000**		4	0.813	0.124	.000**	
			.278	.028	.000**		Sentence Length	1	3	1.487	0.431
.084	.024	.004*	4	4.258	0.472	.000**					
Adjective Variation	1	4	.024	.007	.005*	2		3	2.640	0.444	.000**
			.025	.007	.005*	4		5.411	0.485	.000**	
Adverb Variation	1	3	.022	.007	.006*	4	3	-2.771	0.409	.000**	
			.037	.007	.000**	2	3	.023	.007	.005*	
	.038	.007	.000**	4	.038			.007	.000**		

Note: Textbook-Based = 1, Test-Based = 2, SNS Posts = 3, SNS Comments = 4; Bonferroni adjustment, .05/8 = .0063*, .01/8 = .0013**

Results from answering research question two explored similarities and differences in syntactic and lexical complexity among the different writing genres. Facebook posts and comments resulted in fewer instances of awkward sentences and misused words and this was attributed to shorter

sentence length and simpler vocabulary choice associated with the genre of SNS writing.

Topics for Facebook discussion are near limitless; suggesting a wider variety of vocabulary would be practiced. Even greater differences may exist between test-based writing and weekly class assignment writing tasks (e.g., textbook-based or SNS-based writing). The test-based corpus in the current study came from university entrance exam questions. Academic stakes are greater with test-based writing, and this may result in differences in writing complexity which could lead to errors related to sentence and word complexity as more advanced structures are attempted.

Pedagogical Implications

Certain L2 writing mistakes were constant throughout each genre but unique differences did exist and this leads us to make targeted pedagogical recommendations. Students in the test-based corpus group made considerably more grammar and wrong-word choice mistakes than in the textbook-based and SNS-based groups, possibly due to lack of time afforded through asynchronous writing. To help students practice for high stakes writing tests, we recommend students keep an archive of automatically generated feedback so they can practice metacognitive strategies, like reflection, to decrease common errors over time. Clarity issues with high frequency errors must be achieved before test raters can consider factors related to vocabulary and sentence complexity. Overcoming high frequency mistakes with article usage can be achieved with consistent corrective feedback. The test-based group committed twice as many determiner errors related to article usage so this is an area that should be addressed early with test preparation programs that incorporate AWE.

The textbook-based group fell in between test-based and SNS-based writing for many of the Grammarly metrics observed. The extra time and process writing instruction resulted in relatively equal sentence complexity with test-based writing but fewer errors. Low word variation echoed what was observed with the test-based group and this was partly attributed to students working off of textbook writing models. The same students borrowed language from the same textbook examples just as the same students in the testing condition borrowed language from the essay test questions. Contrarily, students in the SNS-based group were responsible for developing their own writing topics, and this resulted in a broader range of vocabulary use. We recommend students use automatic writing aides like Grammarly to identify overly used words and replace them with synonyms. Non-native English-speaking university students require paraphrasing skills. Several scholars (Leask, 2006; Liu, 2005) have raised concerns of plagiarism among different cultural groups highlighting a need for paraphrasing skills among L2 English speaking university students. Novice writers rely on source text language extensively while more advanced ones are capable of paraphrasing (Keck, 2014). Practicing paraphrasing skills early on with one's own writing may better equip students for paraphrasing in academic or business writing. A final recommendation for the use of AWE with process writing instruction pertains to the timing of AWE implementation. Grammarly may be most valuable for L2 writers immediately after composing first drafts. With immediate feedback, students can recognize possible areas of concern before sending their writing to the instructor.

The SNS-based writing corpus revealed stark differences with test-based and textbook-based writing, especially concerning word variation, sentence complexity, and error frequency. SNS-based writing emanating from the corpus analyzed in the current study can best be likened to shared class diaries of daily activities and personal interests (Bailey et al., 2017). The increased word variation is ascribed to the idiosyncrasies among the Facebook group members' interests and daily events. Students contributing to SNS discussion forums are left to their own free will to choose what topics will be discussed, and this created real world opportunity to use a wide range of vocabulary. Word variation and idea-generation was further amplified with the use of multi-modal communication common in online discussion threads. Over 95 percent of Facebook posts were in reference to images, and many of the images were accompanied by hyperlinks, emoticons, or videos.

Instructors are recommended to host SNS forums for students learning to use new vocabulary in authentic settings. The SNS group used simpler sentence structure and vocabulary than the other groups, and this was especially true for SNS comments and replies. The lower stakes involved with SNS writing meant that the students had no need to impress their instructors or test raters. The SBS activity was graded as pass/fail with main criteria being word count and the number of contributions (i.e., posts and comments). Students had nothing to gain by using more sophisticated vocabulary or sentences and this resulted in fewer accuracy issues and consequently greater clarity. Students were sharing their writing with classmates so they may have invested more energy on accuracy and clarity to avoid looking foolish (Krashen, 1985).

Students in the SNS group had model examples of good writing. Role model examples of good posts were provided by L2 proficient writers and this acted as samples for less capable writers to follow. An advanced writer may write, "Today I went to the *beach* with my *friends* and afterward we had *seafood* for *dinner*." A less L2 proficient classmate can use this sentence as a template to write something like, "This *morning* I went to *church* with my *family* and afterward we had *pizza* for *lunch*." Bailey et al. (2017) recognized that students were borrowing language from one another and at the same time practicing new vocabulary in a real world context. Peer modeling is a characteristic of collaborative learning that increases self-efficacy beliefs (Bandura, 1996), and role model posts on SNS helped others produce compositions with fewer accuracy mistakes. In addition to borrowing language from their classmates' posts, students could practice recursive writing strategies (e.g., second drafts) by borrowing language from their own past posts. A student can use similar language to describe both a dinner with friends and a lunch with family. After a few posts with similar communication goals, the accuracy improves because students are following similar sentence patterns and using similar adjectives but for different communication purposes.

AWE implementation is not a replacement for traditional writing strategies but instead an addition to existing ones. Grammarly should be added to the language learner's L2 writing strategy repertoire.

Writing strategies occur at different stages of the writing process. We recommend Grammarly's use during the when-writing and post-writing stages. Planning, brainstorming, and

outlining should be carried out without concern of writing accuracy. L2 students are even encouraged to use their L1 during the pre-writing stage for idea-generation. Ericsson (2006) recommends automatically generated feedback for writing assistance rather than writing assessment, and the strategy of using AWE as a personal writing assistant is what makes platforms like Grammarly so popular. Successful language learners use a wider array of language learning strategies than their less successful counterparts (Oxford, 1990), suggesting AWE be an addition to the learner's strategy toolbox.

Conclusion

Results from this study provide insight into the types of errors Grammarly recognizes among conceptually unique genres of L2 writing. Instructors of English as a second or foreign language can apply findings in the present study to support the development of their students' writing skills. Students using AWE can save time and increase their confidence when writing in a second language because of fewer accuracy mistakes, and instructors can focus their corrective feedback on higher order writing issues related to rhetoric, tone, cohesion, and style.

Several themes were recognized among error types, error frequency, and writing complexity for the test-based, textbook-based, and SNS-based corpora. Grammarly was successful at identifying local level errors in L2 writing regardless of the writing genre. However, distinct differences were recognized. High stakes testing resulted in more risk taking with vocabulary and sentence complexity which came at the cost of readability (i.e., clarity). Variation errors are mistakes due to repeating the same words frequently. Such vocabulary repetition was common with test-based writing, indicating a need to use AWE programs like Grammarly for paraphrasing practice. Grammarly was most successful at identifying and correcting determiner errors which proves valuable for East Asian students because such errors are common with Chinese, Korean, and Japanese L2 English writers.

Further errors recognized were related to spelling, punctuation, and wordiness. In general, Grammarly incrementally improved writing compositions. A low- quality composition cannot be transitioned into a high quality of writing with Grammarly alone, but instead writing quality improves at incremental levels.

A few limitations with this study are worth mentioning. Findings may not generalize outside the Sino-Tibetan-Austronesian family of languages. Writing samples were collected from intermediate L2 proficiency students. Error types and error frequency would be different depending on the writers' L2 proficiency levels (Liardét, 2015). Finally, results may be different depending on which AWE programs is being used. For instance, Criterion (www.criterion.ets.org) and Virtual Writing Tutor (www.virtualwritingtutor.com) are more focused on giving feedback to students in an academic setting while Grammarly targets users who are both both L1 and L2 students or working professionals.

Future AWE research should explore the effect of integrating tools like Grammarly into the

L2 writing process. AWE technology should not replace existing L1 or L2 writing strategies but instead complement them. Future study may wish to investigate how AWE platforms can work in concert with one another.

Moving forward, EFL/ESL educators should consider how best to implement AWE technology into their writing and communication programs. There is no substitute for human feedback when it comes to the nuances of global level writing errors related to meaning-making and cohesion but such mistakes cannot be addressed when instructors are overwhelmed with local level, treatable, errors.

AWE tools like Grammarly increase writing fluency by saving time composing ideas and this equates to greater language output but at what cost? The ultimate goal should be to help the writer become self-reliant. There is a need now to better understand how dependency on writing aid technology influences self-reliance.

References

- Anson, C. (2006). Can't touch this: Reflections on the servitude of computers as readers. In P. F. Ericsson, R. Haswell (eds), *Machine scoring of human essays* (pp. 38–56). Logan, UT: Utah State University Press.
- Bailey, D. B., Park, I.-W., Haji, S. A. (2017). An investigation of Facebook for language learning: Better understanding perceptions and participation. *CALL-EJ*, 18(2), 12-28. Retrieved from <http://callej.org/journal/18-2/Bailey-Park-Haji2017.pdf>
- Burston, J. (2008). Review of BonPatron: An online spelling, grammar, and expression checker. *CALICO Journal*, 25(2), 337-347. doi:10.1558/cj.v25i2.337-347
- Cavaleri, M., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, 10(1), 223-236. Retrieved from <https://journal.aall.org.au/index.php/jall/issue/view/22>
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267-296. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)
- Chen, C.-F., & Cheng, W.-Y. (2008). Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning and Technology*, 12(2), 94-112. Retrieved from <https://www.lltjournal.org/item/2631>
- Chung, G. K. W. K., & Baker, E. L. (2003). Issues in the reliability and validity of automated scoring of constructed responses. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 23-40). Mahwah, NJ: Lawrence Erlbaum.
- Dikli, S. (2010). The nature of automated essay feedback. *CALICO Journal*, 28(1), 99-134. Retrieved from <https://www.jstor.org/stable/calicojournal.28.1.99>
- El Ebyary, K., & Windeatt, S. (2010). The impact of computer-based feedback on students' written work. *International Journal of English Studies*, 10(2), 121-142. Retrieved from <https://files.eric.ed.gov/fulltext/EJ936915.pdf>
- Ericsson, P. F. (2006). The meaning of meaning: Is a paragraph more than an equation. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays* (pp. 28-37). Logan, UT: Utah State University Press.
- Ferris, D. R. (2004). The “grammar correction” debate in L2 writing: Where are we, and where do we go from here? *Journal of Second Language Writing*, 13(1), 49-62. <https://doi.org/10.1016/j.jslw.2004.04.005>
- Ferris, D. R., Chaney, S. J., Komura, K., Roberts, B. J., & McKee, S. (2000). Perspectives, problems, and practices in treating written error. In Colloquium presented at International TESOL Convention, Vancouver, B.C., March 14-18, 2000.
- Gauthier, M. (2013). Anglophone high school boys' engagement and achievement in editing their French writing using the Bon Patron Pro. *Journal of Classroom Research in Literacy*, 6, 24-35. Retrieved from <https://jcr.l.library.utoronto.ca/index.php/jcr/article/view/16093>

- Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Language, and Assessment*, 8(6), 1-43. Retrieved from <https://ejournals.bc.edu/index.php/jtla/article/view/1625>
- Halliday, M. A. K. (1989). *Spoken and Written Language*. Oxford: Oxford University Press.
- Hamilton, C. E. (2015). The contributions of systematic functional grammar to the error analysis framework. *TESOL International Journal*, 10(1), 11-28. Retrieved from <http://tesol-international-journal.com/112/volume-10-issue-1-2015/>
- Han, N.-R., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, 12(2), 115-129. doi:10.1017/S1351324906004190
- Hoang, G. T. L., & Kunnan, A. J. (2016). Automated essay evaluation for English language learners: A case study of MY Access. *Language Assessment Quarterly*, 13(4), 359–376. <https://doi.org/10.1080/15434303.2016.1230121>
- Katayama, A. (2007). Japanese EFL students preferences towards correction of classroom oral errors. *The Asian EFL Journal*, 9(4), 289-305. Retrieved from <https://www.asian-efl-journal.com/main-journals/japanese-efl-students-preferences-toward-correction-of-classroom-oral-errors/>
- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practice. *Journal of Second Language Writing*, 25, 4-22. <https://doi.org/10.1016/j.jslw.2014.05.005>
- Kelly, C., & Gargagliano, A. (2011). *Writing From Within 2*. Cambridge, UK: Cambridge University Press.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. New York: Longman.
- Leask, B. (2006). Plagiarism, cultural diversity and metaphor—implications for academic staff development. *Assessment & Evaluation in Higher Education*, 31, 183–199. <https://doi.org/10.1080/02602930500262486>
- Liardét, C. (2015). Academic literacy and grammatical metaphor: Mapping development. *TESOL International Journal*, 10(1), 29-46. Retrieved from <http://tesol-international-journal.com/112/volume-10-issue-1-2015/>
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27(1), 1-18. <https://doi.org/10.1016/j.jslw.2014.10.004>
- Liao, H. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System*, 62(1), 77–92. <https://doi.org/10.1016/j.system.2016.02.007>
- Liu, D. (2005). Plagiarism in ESOL students: Is cultural conditioning truly the culprit? *ELT Journal*, 59, 234–243. <https://doi.org/10.1093/elt/cci043>
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners’ oral narratives. *The Modern Language Journal*, 96(2), 190-208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

- McGee, T. (2006). Taking a spin on the Intelligent Essay Assessor. In P. F. Ericsson & R. H. Haswell (Eds.), *Machine scoring of student essays* (pp. 79-92). Logan, UT: Utah State University Press.
- Nadasdi, T., & Sinclair, S. (2007). Anything I can do, CPU can do better: A comparison of human and computer grammar correction for L2 writing using BonPatron.com. Retrieved from <http://bonpatron.com>
- Newgarden, K. (2009). Annotated bibliography: Twitter, social networking and communities of practice. *TESL-EJ*, 13(2), 2-13. Retrieved from <http://www.tesl-ej.org/pdf/ej50/m2.pdf>
- O'Neill, R., & Russell, A. M. T. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1), 42-56. <https://doi.org/10.14742/ajet.3795>
- Potter, R., & Fuller, D. (2008). My new teaching partner? Using the grammar checker in writing instruction. *English Journal*, 98(1), 36-41. Retrieved from <https://www.jstor.org/stable/40503205?seq=1>
- Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge UP.
- Stevenson, M. (2016). A critical interpretive synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition*, 42(1), 1-16. <https://doi.org/10.1016/j.compcom.2016.05.001>
- Stevenson, M., & Phakiti, A. (2014). The effect of computer-generated feedback on the quality of writing. *Assessing Writing*, 19(1), 51-65. <https://doi.org/10.1016/j.asw.2013.11.007>
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Eds.), *Sociocultural theory and second language learning* (pp. 97-114). London: Oxford University Press.
- Ware, P. (2011). Computer-generated feedback on student writing. *Tesol Quarterly*, 45(4), 769-774. <https://doi.org/10.5054/tq.2010.272524>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36(1), 90-102. <https://doi.org/10.1016/j.asw.2018.02.004>