UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

http://revistas.um.es/ijes

AWE for Formative Purposes in the EFL Context: A Study of Multiple Revisions in Academic Abstract Writing

MANUEL RODRÍGUEZ-PEÑARROJA*
Universitat Jaume I (Spain)

Received: 10/04/2024. Accepted: 09/01/2025.

ABSTRACT
The use of automated writing evaluation (AWE) tools for writing practices has become a central issue in English as a second and foreign language teaching contexts. Researchers appear to agree that students' positive outcomes hinge upon their integration with formative purposes. Nonetheless, few studies have addressed multiple revisions and automated feedback provision in students' academic writing development. In this context, an instructional treatment for the integration of this technology has been designed. Explicit instruction on academic writing, AWE workshops, and practice activities remain at its core. Hence, this paper examines the effects of self- and AWE-mediated writing revisions on undergraduate learners' academic writing performance over time and at different levels of proficiency. A series of repeated-measures analysis of variance (ANOVA) were computed to study participants' syntactic complexity, readability and language issues, and lexical diversity outcomes. Results revealed improvement in some of the dependent variables, that is, language issues reduction and increased type-token ratio mean scores, which could represent an initial step towards reconsidering the provision of automated written feedback.

KEYWORDS: academic writing; automated writing evaluation feedback; English as a foreign language; explicit instruction; writing revision.

1. INTRODUCTION

Learning to write in English in second language (L2) and foreign language (EFL) learning contexts has been perceived as an arduous task on the part of students (Alameddine & Mirza, 2016; Muñoz-Luna, 2015). This is on the grounds that the processes involved in writing skill

Address for correspondence: Office HC0314DD, Facultad de Ciencias Humanas y Sociales. Universitat Jaume I; Av. Vicent Sos Baynat, s/n 12071 Castellón de la Plana, España; e-mail: penarroj@.uji.es

Print ISSN: 1578-7044; Online ISSN: 1989-6131
doi: 10.6018/ijes.611761

development challenge learners' comprehension, socio-affective, cognitive, and metacognitive strategies in conjunction with textual, lexicogrammatical, and discursive demands (Lillis & Curry, 2006). Recent advancements in information and communication technology (ICT) have given rise to automated written evaluation (AWE) systems designed for the computerized analysis and feedback provision of written texts (e.g., Grammarly). Over the last two decades, language instructors and researchers have focused on these tools' effects on students' writing progression, feedback engagement, and language proficiency (Huang & Renandya, 2020; Koltovskaia, 2020; Lee, 2020; Li, 2021; Link et al., 2020; Zhang, 2020). Dissimilarities in findings, the acknowledged importance of writing revision (Barkaoui, 2016; Chen & Zhang, 2019), and the need for instructing L2 learners in the academic writing register in higher education (HE) contexts have motivated the corresponding study.

Grounded in the advantages of explicit instruction and enhanced practice opportunities for writing improvement (Kobayashi & Rinnert, 2001; Sengupta, 2000), this paper proposes an instructional treatment for introducing undergraduate university students to the basic aspects of academic writing and AWE tools usage. The objective is to delve into the effect of explicit-deductive teaching, practice activities, and multiple self- and AWE-mediated writing revisions on EFL students' written outcomes. Innovation resides in the number and type of revisions carried out by students and the use of different AWE online tools. Likewise, this research aims to respond to the lack of studies on the use of AWE in the academic writing register (Guo et al., 2021; Strobl et al., 2019) and the prospects for integrating AWE tools into instructional practices (Chapelle et al., 2015; Li et al., 2017). Participants' proficiency is also examined as mediating AWE feedback understanding, engagement, and uptake (Guo et al., 2021, Koltovskaia, 2020, Link et al., 2020; Zheng & Yu, 2018).

2. LITERATURE REVIEW

2.1. Academic writing in HE contexts

Biber et al. (2002) define academic writing as an expository register with its main communicative purpose being to inform a specialized audience through empirical and theoretical argumentation adhering to academic writing conventions. Accordingly, L2 students experience an increase in writing demands as they gradually master their writing skills at different learning stages and levels of language proficiency (Council of Europe, 2020). Thus, the progression from writing topic-based descriptive texts, reports, and argumentative essays to academic assignments calls for students to adopt a more creative, critical, and analytical writing style in the university context. This is on the grounds that L2 learners' previous writing experiences are limited to writing topic-based content as required in a specific essay typology (Spring et al., 2010). Even though the structuring of information

Print ISSN: 1578-7044; Online ISSN: 1989-6131

into pre-patterned paragraphs allows the writer less freedom, creativity, and self-reasoning (Davidson, 2019), the convenience of rhetorical patterns to integrate lexicogrammatical, coherence, and cohesion resources at certain writing stages must not be neglected. Hence, L2 students may struggle with the merging of their previous writing background as they progress to producing academic texts.

Advancements in ICT have spawned new writing practices, changing the nature of this skill to a certain extent (Li et al., 2017). Students finding ICT easy to use, and its positive effects on motivation and task accomplishment (Kaharuddin, 2020), contrast with the still-under-debate integration of their digital competence in the educational context (Waycott et al., 2010). Nevertheless, motivated by a steadily increasing digitization of education, the use of ICT for writing purposes has become widespread among students and teachers worldwide (Guo et al., 2021). Li et al. (2017) identify a series of technologies for L2 writing and classify them into Web 2.0 applications, corpus-based tools, and AWE systems, the last being the most widely researched in the recent literature (Strobl et al., 2019; Thi & Nikolov, 2023; Wilson & Roscoe, 2020; Xu & Zhang, 2022) on the central topic of this study.

2.2. AWE tools: initial considerations

Also termed automated written corrective feedback (AWCF) (Ranalli, 2018), AWE tools "extract linguistic, syntactic, semantic, or rhetorical features of text related to writing quality" (Wilson & Roscoe, 2020, p. 88), score written texts using machine-learning algorithms, and provide corrective, formative, and diagnostic feedback for self-revision (McCarthy et al., 2022). Corrective feedback entails the identification of language issues and subsequent direct corrections or alternatives to choose from. Formative feedback, usually presented as report scores, equips students with information "about their actual state of learning performance in order to modify the learner's thinking or behaviour in the direction of learning standards" (Goldin et al., 2017, pp. 385–386). Diagnostic feedback comprises students' awareness on specific writing issues and is aimed at helping them learn through guided revisions (Chapelle et al., 2015).

Strobl et al. (2019) identify 44 tools (e.g., Criterion, My Access, PEG Writing) to support writers in HE contexts. Given their customization differences and the types of feedback provided (Cotos, 2023), their implementation in EFL–ESL instructional settings depends primarily on the purposes for their application (Hockly, 2019; Rodríguez-Peñarroja, 2022). The use of these systems for summative assessment entails the computerized grading of students' written assignments based on the scores provided by these tools. Despite their convenience as time-saving systems (Ranalli, 2018), "the ability of AWE software to judge critical thinking, rhetorical knowledge, creativity, or a student's ability to tailor their text to a specific readership" (Hockly, 2019, p. 83) is still a bone of contention (Herrington & Moran, 2012; Stevenson & Phakiti, 2014; Zhang, 2020). AWE tools use for formative purposes

Print ISSN: 1578-7044; Online ISSN: 1989-6131

encompasses students' writing progress through computer-based feedback provision, as these systems flag and describe a wide typology of lexicogrammatical issues at sentence and paragraph levels. This practice has proved positive for writing revisions (Chapelle & Voss, 2017; Huang & Renandya, 2020; Li et al., 2015; Thi & Nikolov, 2023; Warschauer & Ware, 2006) as well as enhancing students' motivation and self-efficacy (Wilson & Roscoe, 2020).

In this light, Chappelle et al. (2015) and Cotos (2014, 2023) appraise the formative use of AWE tools as being subject to their integration into instructional practices. The implementation of machine feedback has been found to depend on its accuracy and length (Koltovskaia, 2020; Ranalli, 2018), as well as on students' L2 proficiency (Thi & Nikolov, 2023) and their engagement with the feedback provided (Huang & Renandya, 2020; Liao, 2016; Zheng & Yu, 2018). On the topic of proficiency, previous studies' results are often contradictory. Ranalli (2018) suggests that AWE tools' provision of surface-level feedback may result in low-proficiency students' gains in writing accuracy, since low-level (LL) writing aspects (e.g., grammar, spelling, and punctuation) can be self-addressed using these technologies. Conversely, Zheng and Yu (2018) and Xu and Zhang (2022) note that these students' understanding of feedback provision may be limited by their linguistic competence. Thus, they might not benefit from AWE feedback provision and require additional support from language instructors. Concerning high-proficiency students, they may fail to engage with AWE feedback since some high-level (HL) issues (e.g., style, content, information structuring, cohesion, and coherence) (Link et al., 2020) are not diagnosed by these tools (Chen & Cheng, 2008).

2.3. AWE revision operations in the ESL–EFL context

Empirical evidence supports the importance of revising written texts (Barkaoui, 2016; Chen & Zang, 2019; Hayes, 2012), as the lack of revision remains a significant challenge for L2 writers (Chambers, 2011). The process of detecting, diagnosing, and addressing language and mechanical issues in writing relies on students' ability to engage in effective revisions. Research has shown that explicit instruction aimed at improving revision skills is essential (Kobayashi & Rinnert, 2001; Sengupta, 2000), particularly because L2 writers tend to revise their texts more frequently (Stevenson et al., 2006). Whereas more competent writers show concern with HL revisions, their less experienced counterparts concentrate on LL aspects (Koltovskaia, 2020).

In the recent literature, several studies have pointed towards the convenience of using AWCF in ESL–EFL teaching and learning contexts, since these tools yield objective writing scores, instant corrective feedback, and revision opportunities. Moreover, students' motivation and autonomy for drafting, revising, redrafting, and polishing their written assignments can be fostered (Dikli, 2006; Li et al., 2015; Liao, 2016; Wang et al., 2013). Teachers may also benefit from their use, as they can spend more time addressing the aims of

Print ISSN: 1578-7044; Online ISSN: 1989-6131

writing instruction and HL writing concerns (Link et al., 2020; Ranalli, 2018). Some of the drawbacks encompass the complexity and extensiveness of the AWE feedback provided. This may hinder its understanding and limit writers' corrections to LL concerns (Dikli, 2006; Jingxin & Razali, 2020; Ranalli, 2018; Wang et al., 2013) leaving aside HL requirements and social and contextual aspects (Zhang, 2020). AWE feedback provision has also been found to misidentify language issues (Chapelle et al., 2015; Guo et al., 2021), which may directly affect students' engagement and the correctness of a final piece of writing.

Lee (2020) examined the effects of the AWE tool Criterion over a year in relation to participants' writing improvement, overall English proficiency development, and perceptions of usage. The results of a test-retest experimental design suggested a decrease in the number of writing mistakes, along with improvements in style and grammar, as well as increases in word count, T-Units, and confidence. Zhang's (2020) study delved into students' perceptions and feedback engagement with the tool Pigai. Findings indicate EFL participants' engagement and positive attitudes towards the feedback provided; however, revisions were mainly restricted to LL concerns. Similarly, Huang and Renandya (2020) explored Pigai's AWE feedback provision effects on EFL low-proficiency undergraduate students. The pre- and post-test comparison revealed that participants perceived the software as useful for enhancing their writing performance, although they did not report noticeable improvement. Such inconclusiveness has given rise to investigations into feedback provision and combination. Some studies advocate for the positive long-term effect of combining AWE and instructor feedback provision (Lee et al., 2013; Liao, 2016; Link et al., 2020), whereas Li (2021) contends that combining feedback can be fruitless in reducing grammar mistakes.

Regarding the written genres researched, Strobl et al. (2019) and Guo et al. (2021) call for the need to investigate the effects of AWE use in academic writing, an underexplored register in AWE research literature. Guo et al. (2021) examine the effects of using Grammarly among thirty-six doctoral and undergraduate students enrolled in two academic research writing courses. The authors explore the degree of Grammarly feedback effectiveness in reducing errors and other factors related to students' revisions and responses to feedback. These are the acceptance, substitution, and rejection of grammar, spelling, punctuation, and writing conventions based on AWE-flagged errors and accuracy suggestions. The experimental test–retest design combined the writing of a first draft of the introduction of a research paper with the revision of a second draft, using this tool. Results showed a decrease in Grammarly-flagged errors, with students correctly addressing 85% of flagged usages. Concerning the adoption of feedback, results showed that revisions, on average, indicated 58% acceptance and 29% rejection, with 13% being students' own corrections. The authors conclude that students should learn to use AWCF critically by exercising judgement about the flagged errors and accuracy suggestions.

Therefore, this paper seeks to address the effect of multiple self- and AWE-mediated revisions on undergraduate EFL students' academic writing performance. This considers i)

Print ISSN: 1578-7044; Online ISSN: 1989-6131

the scarcity of studies on the use of AWE in the academic writing register as outlined by Guo et al. (2021) and Strobl et al. (2019); ii) the wide agreement that AWE tools with formative purposes should be integrated in instructional practices for a wider understanding of their usability and limitations (Chapelle et al., 2015; Cotos, 2014; Li et al., 2017; Rodríguez-Peñarroja, 2022); iii) the importance of writing revisions (Barkaoui, 2016; Chambers, 2011; Chen & Zhang, 2019; Hayes, 2012); and iv) the benefits of explicit instruction to enhance writing improvement (Kobayashi & Rinnert, 2001; Sengupta, 2000). The main objective is to explore multiple revisions' effects on participants' written performance at different levels of proficiency, participants having received prior instruction on using AWE tools and academic abstract writing. The following research questions were formulated:

-How do explicit instruction and the use of AWE tools impact EFL learners' academic writing performance, particularly through self- and AWE-mediated revisions?
-What is the effect of participants' English proficiency on self- and AWE-mediated revisions over time?

3. METHODOLOGY

3.1. Participants and context

A total of 146 undergraduate EFL students enrolled in the English Studies degree at a university in Spain participated in the study. Plagiarism rates, writing task incompleteness, and absences in the academic writing seminars and AWE workshops limited the final pool of participants to 110, who reported an age range of 19–23 (M = 19.87, SD = .92). Prior to the instructional treatment, participants' English proficiency was measured with the Oxford University Press quick placement test version 1 (UCLES, 2001); see Table 1.

Table 1. Participants' proficiency frequencies (CEFR).

		Range values	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	B1	30–39	29	26.4	26.4	26.4
	B2	40–47	40	36.4	36.4	62.7
	C1	48–54	41	37.3	37.3	100.0
	Total		110	100.0	100.0	

3.2. Measures and instruments for data collection and analysis

To investigate the impact of self- and AWE-mediated multiple revisions on students' academic abstract writing, syntactic complexity, readability and language issues, and lexical

Print ISSN: 1578-7044; Online ISSN: 1989-6131

diversity have been studied. The analysis of syntactic complexity, referred to as "the sophistication of syntactic forms produced by a speaker or writer and the range or variety of syntactic forms produced" (Crossley & McNamara, 2014, p.67), was performed using the syntactic analysis tool TAASSC v.1.3.8 (Kyle, 2016). This consisted of eight variables from the L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010), which have been informed as L2 syntactic complexity reliable indicators including global, phrasal, and clausal measures (Casal & Lee, 2019; Eckstein & Bell, 2023; Norris & Ortega, 2009). The measures include the mean length of clause, sentence, and T-unit (MLC, MLS, and MLT, respectively), with the T-unit complexity ratio (C/T) and the dependent clause ratio (DC/C) as subordination measures, and two measures for complex nominals: the complex nominals per clause (CN/C) and complex nominals per T-unit (CN/T).

Readability measures include the Flesch reading ease (FRE) (Flesch, 1979) and the Flesch-Kincaid grade level (FKGL). The FRE measures the difficulty of understanding reading content in English; it is calculated taking into consideration the ratio of total words, syllables, number of sentences, and the number of words per sentence. The resulting values are codified in a 0–100 scale; the easier a text is to read, the higher it is scored, while more complex texts present lower scores (Corazza et al., 2018). The FKGL is aimed at determining the education level required for readers to read and understand a text (Eleyan et al., 2020). While using the same measurements as FRE but different weighting factors, a text scored from 0 to 30 is understood by university graduates demonstrating higher writing complexity, while a 90 to 100 scored text is understood by average 11-year-old students. The analysis of language issues (LI) was performed using the CorrectEnglish online tool that flags improper language use (e.g., colloquialisms, capitalization, spelling, punctuation, word misuse, etc.). Written outcome variability in FRE, FKGL, and LI was examined to test participants' abstracts readability and linguistic accuracy progress.

A reliable indicator of writing proficiency (Gregori-Signes & Clavel-Arroitia, 2015; Johansson, 2008), lexical diversity refers to the number of lexical words (i.e., nouns, adjectives, verbs, and adverbs) per total words or tokens, that is, the type-token ratio (TTR). Noted by Nasseri and Thompson (2021, p. 1), "a dense use of nouns, is regarded as an indicator of condensed academic writing and advanced information prose, and as a strong predictor of academic writing proficiency". The Coh-Metrix tool (McNamara et al., 2014) measured the readability and lexical diversity (i.e., tokens, types, and the TTR) of abstracts over revisions. The institutional Moodle platform was used for the uploading of abstract revisions, and statistical analyses have been computed on SPSS v.26.

3.3. AWE tools for writing revision

ProWritingAid and Grammarly have been selected as the online tools for abstract revision, since these allow users to set the register to academic writing and have been widely studied in

Print ISSN: 1578-7044; Online ISSN: 1989-6131

the literature (Dizon & Gayed, 2021; Koltovskaia, 2020; Nova, 2018; Soleimani & Moqimi, 2023). ProWritingAid inspects paragraph readability, sentence length, and structure, and generates a series of reports for self-revision. A short description of the issues found and suggestions for improving and fixing language problems are provided, therefore not limiting the given feedback to merely corrective. The "core" report brings into focus the use of grammar and writing style. The "repeats" report includes the identification of repeated words and word phrases. The "structure" report presents issues regarding sentence structure, length, and in-between sentences transition. The "readability" report identifies jargon words, vocabulary use, and sentence length where this may affect readability. Likewise, Grammarly can be customized in relation to the audience, degree of formality, domain, writing tone, and writers' communicative intention. This tool scores written texts and identifies plausible text improvements, providing short descriptions of the issues identified and more suitable alternatives. The four areas analysed are correctness, clarity, engagement, and delivery, i.e., writing consistency with respect to the type of preselected audience.

3.4. Instructional treatment

Students must write a 200–350-word research abstract based on a published research paper of their choice as part of their continuous assessment. The writing task has been designed and sequenced to cover the whole semester. However, mandatory curricular needs limited the instructional treatment to two seminars on academic register and abstract writing and two workshops on the use of Grammarly and ProWritingAid AWE tools (Table 2).

Table 2. Instructional treatment – outline.

Week	Sessions	Treatment	Objectives
1–2	2 × 120 min. Seminars 1 and 2	Explicit instruction Practice activities	Describe individual project (goals and deadlines) Raise awareness of academic writing importance Familiarize students with academic writing basic aspects Introduce students to abstract writing
3–7	Students' own work	--	Search and select a research paper Practice academic register reading and comprehension skills Practice academic writing and information synthesis skills Produce a first draft abstract version
8–9	2 × 120 min. AWE workshops	Explicit instruction Abstract revision	Promote students' digital competence Promote self-revision practices Produce a first self-revised abstract version (T1) Enhance abstract revision through AWE use Raise critical assessment of the feedback provided Produce 2 AWE-revised versions (T2 and T3)
10–15	Students' own work	--	Practice revision skills and critical self-assessment of T1, T2, and T3 drafts Write a final abstract version (T4)

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Weeks 1–2 present the individual project to students. The first seminar consists of a brief description of the general aspects of academic writing and an introduction to the main traits of a formal writing style. The second seminar focuses on abstract writing as the short summary of a whole research paper. The parts of a complete abstract are described, exemplified, and analysed in published research papers. Active practice, engaging in the identification and ordering of abstract parts along with rewriting activities on the use of specific phraseology as found in Wallwork (2011), follows theoretical explanations and examples. In weeks 3–7, students work autonomously to search for an empirical research paper and write an alternative abstract. Plagiarism rates and textual similarity are controlled using the Turnitin plagiarism detection tool.

Weeks 8 and 9 are used to guide students through multiple revisions. In week 8, participants self-revise the initial version of their abstract by checking its parts, content, lexicogrammatical aspects, and textual coherence and cohesion. This initial self-revised draft (T1) is uploaded to Moodle, followed by in-class workshops using AWE for T1 revisions. First, ProWritingAid settings are customized by selecting British English and the academic writing document type. Then, the instructor presents the available feedback reports (i.e., core, repeats, structure, and readability) and shows students how to apply the suggested changes if necessary. Having revised their abstracts, students upload their second draft version (T2) to Moodle. In week 9, Grammarly is set up, addressing a knowledgeable audience, with a formal writing tone, academic domain, and informative and descriptive communicative intent. Students copy and paste their T1 for a second AWE-mediated revision, following the instructor's description of the feedback provided in the clarity, engagement, and delivery reports. Having applied the necessary changes, students upload their third draft (T3). Weeks 10–15 are devoted to students' writing of a final version of their abstract (T4).

Participants' T1 to T4 non-linear revision was a methodological decision on the grounds that ProWritingAid and Grammarly are based on different models of generative artificial intelligence. Hence, computerized feedback reports may differ in the number and type of flagged language issues and suggestions for writing improvement at LL and HL. This may allow participants to produce a final version (T4) of their abstract, having considered their initial draft (T1) and the feedback on the two revised and polished versions (T2 and T3).

4. RESULTS

The results section is divided into two sub-sections, which include the main aspects under analysis, i.e., syntactic complexity, readability and language issues, and lexical diversity. To assess the impact of the devised instructional treatment and of multiple draft revisions on students' academic writing performance (RQ1), a repeated-measures ANOVA with Bonferroni post-hoc tests with thirteen within-subjects dependent variables at four time

Print ISSN: 1578-7044; Online ISSN: 1989-6131

points (T1–T4) was computed. Then, participants' English proficiency (i.e., B1, B2, and C1) was added to the model as a between-subjects variable to examine its effects on students' written outcomes over time (RQ2).

4.1. RQ1 Instruction effects on participants' written outcomes

4.1.1 Syntactic complexity

Table 3 reveals that participants performed similarly in the *syntactic complexity* measures in the different revisions; nonetheless, mean scores remained higher at T1 revision stage. Results from the repeated-measures analysis of variance (ANOVA) (Table 4) showed significant differences ($p < .05$) across all variables; partial eta squared values were small ($\eta_p^2 \leq .032$), i.e., MLC, C/T, DC/C, CN/C; medium ($\eta_p^2 \leq .14$), i.e., MLT, CN/T; and large ($\eta_p^2 = .146$), i.e., MLS.

Table 3. Descriptive statistics for the syntactic complexity measures over time.

	T1		T2		T3		T4	
Measure	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
MLC	12.72 (2.99)	12.16-13.29	12.35 (2.97)	11.78-12.91	12.55 (2.86)	12.01-13.09	12.59 (2.79)	12.06-13.12
MLS	25.45 (4.36)	24.63-26.28	24.51 (4.35)	23.69-25.34	24.75 (4.39)	23.92-25.58	24.58 (4.43)	23.74-25.42
MLT	23.07 (4.27)	22.26-23.88	22.25 (4.25)	21.44-23.05	22.40 (4.17)	21.61-23.18	22.25 (4.26)	21.44-23.06
C/T	1.86 (.38)	1.79-1.93	1.83 (.39)	1.76-1.91	1.83(.39)	1.76-1.91	1.81(.37)	1.74-1.88
DC/C	.43 (.11)	.41-.45	.42 (.12)	.40-.44	.42(.12)	.40-.44	.41(.11)	.39-.44
CN/C	1.80 (.51)	1.7-1.90	1.76 (.49)	1.67-1.86	1.78(.50)	1.68-1.87	1.77(.48)	1.68-1.87
CN/T	3.27 (.78)	3.12-3.42	3.14 (.75)	2.99-3.28	3.16(.75)	3.02-3.30	3.13(.72)	2.99-3.27

Note: M = mean; SD = standard deviation; CI = confidence interval; MLC = mean length of clause; MLS = mean length of sentence; MLT = mean length of T-unit; C/T = T-unit complexity ratio; DC/C = dependent clause ratio; CN/C complex nominals per clause; CN/T complex nominals per T-unit.

Table 4. Syntactic complexity repeated-measures ANOVA results.

Measure	III sum of square	*df*	*F*	Sig	Partial Eta Squared
MLC	7.411	2.017	3.396	.035*	.031
MLS	62.113	2.755	18.299	.000*	.146
MLT	53.589	2.749	13.835	.000*	.114
C/T	.157	1.984	3.566	.030*	.032
DC/C	.011	2.530	3.447	.023*	.031
CN/C	.113	2.349	3.388	.028*	.031
CN/T	1.439	2.961	12.137	.000*	.102

*. The mean difference is significant at the .05 level.

Note: Partial eta squared values are interpreted as small (.01), medium (.06), or large (.14) effects (Cohen, 1988, pp. 280–287), as discussed in Richardson (2011, p.142).

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Pairwise comparisons contrasting participants' first writing and their subsequent revisions revealed that the number of words per clause (MLC) generally significantly decreased from students' first self-revised draft to T2 and T3 AWE-mediated revisions ($p < .005$). T1–T4 comparison revealed no significant differences ($p > .05$), however. Likewise, the mean length of sentence (MLS) and the number of words per T-unit (MLT) reduced over revisions ($p < .001$). Participants' use of subordination (i.e., C/T and DC/C measures) lessened across all revision stages: this was significant in the number of clauses per T-unit (C/T) T1–T3 ($p < .05$) and in the number of dependent clauses per clause (DC/C) T1–T4 ($p < .05$). The variables measuring the use of complex nominals also were trimmed significantly from T1 to T4. The number of complex nominals per clause (CN/C) was cut down from T1 to T2 and T3 ($p < .05$). Nonetheless, CN/C T1–T4 did not statistically differ ($p > .05$). CN/T measure followed the same decreasing tendency, from T1 to T4 ($p < .001$).

4.1.2. Readability and language issues

Participants showed considerable differences in the reduction of language issues (LI) at the revision stages, whereas the readability measures (FKGL and FRE) presented low variability (Table 5). Results from the repeated-measures ANOVA revealed significant differences in all variables: FKGL $F_{(2.968)} = 17.916$, $p < .01$, $\eta_p^2 = .143$ (large effect size); FRE $F_{(2.701)} = 12.271$, $p < .001$, $\eta_p^2 = .103$ (medium effect size); and LI $F_{(1.996)} = 243.158$, $p < .01$, $\eta_p^2 = .694$ (large effect size).

Table 5. Descriptive statistics for the syntactic complexity measures over time.

	T1		T2		T3		T4	
Measure	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
FKGL	13.44 (2.01)	13.06-13.83	12.88 (1.99)	12.50-13.25	13.15 (2.05)	12.77-13.54	13.16(1.97)	12.78-13.53
FRE	41.22(10.56)	39.22-43.21	42.82(10.48)	40.84-44.80	41.79(10.57)	39.80-43.79	41.59(10.36)	39.63-43.55
LI	7.45(3.31)	6.82-8.07	4.01 (2.66)	3.51-4.51	3.21 (2.49)	2.74-3.68	1.95(1.34)	1.34-2.08

Note: M = mean; SD = standard deviation; CI = confidence interval; FKGL = Flesch-Kincaid grade level; FRE = Flesch reading ease (FRE); LI = language issues.

Table 6. Readability and language issues pairwise comparisons.

						95% CI for Difference[b]	
Measure	(I) Time	(J) Time	DM (I-J)	SE	Sig.[b]	Lower	Upper
FKGL	1	2	.578*	.079	.000	.365	.791
		3	.305*	.074	.000	.105	.505
		4	.308*	.090	.005	.066	.550
FRE	1	2	-1.667*	.313	.000	-2.507	-.826
		3	-.635	.273	.130	-1.368	.097

Print ISSN: 1578-7044; Online ISSN: 1989-6131

		4	-.445	.349	1.000	-1.383	.492
LI	1	2	3.453*	.211	.000	2.885	4.021
		3	4.285*	.264	.000	3.574	4.996
		4	5.777*	.302	.000	4.966	6.588

Based on estimated marginal means.
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

FKGL pairwise comparisons results (Table 6) significantly decreased across revisions ($p \leq .005$). Thus, there was slight improvement in the *FKGL* measure since the T4 mean score prevailed lower over the previous revisions. FRE mean results revealed that more complex texts were achieved in students' T1. As FRE values represent higher complexity at lower mean scores, no progress was identified in the revised abstracts. FRE increased significantly only at T2 revision stage ($p < .001$). The LI gradual decrease was found to be significant across all revision stages ($p < .001$).

4.1.3. Lexical diversity

Table 7 presents lexical diversity descriptive statistics (T1–T4). At the different revision stages, repeated-measures ANOVA showed significant differences for the variables of word tokens $F_{(2.156)} = 7.632$, $p < .01$, $\eta_p^2 = .067$ (medium effect size); word types $F_{(2.536)} = 3.720$, $p < .05$, $\eta_p^2 = .034$ (small effect size), and the TTR $F_{(2.160)} = 3.076$, $p < .05$, $\eta_p^2 = .028$ (small effect size).

Table 7. Descriptive statistics for the lexical diversity measures over time.

	T1		T2		T3		T4	
Measure	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI	M (SD)	95% CI
Tokens	263.57 (36.30)	256.71-270.43	258.55 (34.13)	252.10-265.00	258.61 (36.52)	251.71-265.51	260.00 (32.30)	253.90-266.10
Types	138.80 (19.95)	135.02-142.57	137.32 (19.36)	133.66-140.98	137.80 (19.78)	134.07-141.54	139.10 (18.94)	135.52-142.67
TTR	.53 (.093)	.51-.55	.53 (.094)	.52-.55	.54 (.10)	.52-.56	.54 (.087)	.52-.55

Note: M = mean; SD = standard deviation; CI = confidence interval; TTR = type-token ratio.

Table 8. Lexical diversity pairwise comparisons.

						95% CI for Difference[b]	
Measure	(I) Time	(J) Time	*DM (I-J)*	*SE*	*Sig.*[b]	Lower	Upper
Tokens	1	2	5.071*	1.002	.000	2.379	7.763
		3	5.150*	1.069	.000	2.277	8.023
		4	3.668	1.648	.169	-.762	8.098
Types	1	2	1.417*	.518	.044	.024	2.810
		3	.991	.614	.656	-.660	2.643
		4	-.414	.767	1.000	-2.475	1.648
TTR	1	2	-.005*	.001	.005	-.008	-.001

Print ISSN: 1578-7044; Online ISSN: 1989-6131

3	-.008*	.003	.039	-.017	.000
4	-.008	.003	.107	-.016	.001

Based on estimated marginal means.
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

Pairwise comparisons (Table 8) showed a descending tendency in the number of tokens from T1 to T4, these being significant only at T2 and T3 ($p < .001$). Nonetheless, no statistical differences were found in the T1–T4 comparison ($p = .169$). Likewise, the mean number of word types significantly decreased from T1 to T2 ($p = .044$) revisions, still increasing in T3 and T4. The TTR increased in every revision from T1 to T4. Significant differences were identified at T2 and T3 revision stages ($p < .05$).

4.2. RQ2 Proficiency effects on students' revisions

EFL students' English proficiency was introduced in the model as a between-subjects factor to assess the effect of the instructional treatment and revisions on the dependent variables considering participants' different proficiency levels (Table 9).

Table 9. Tests of between-subject effects for proficiency.

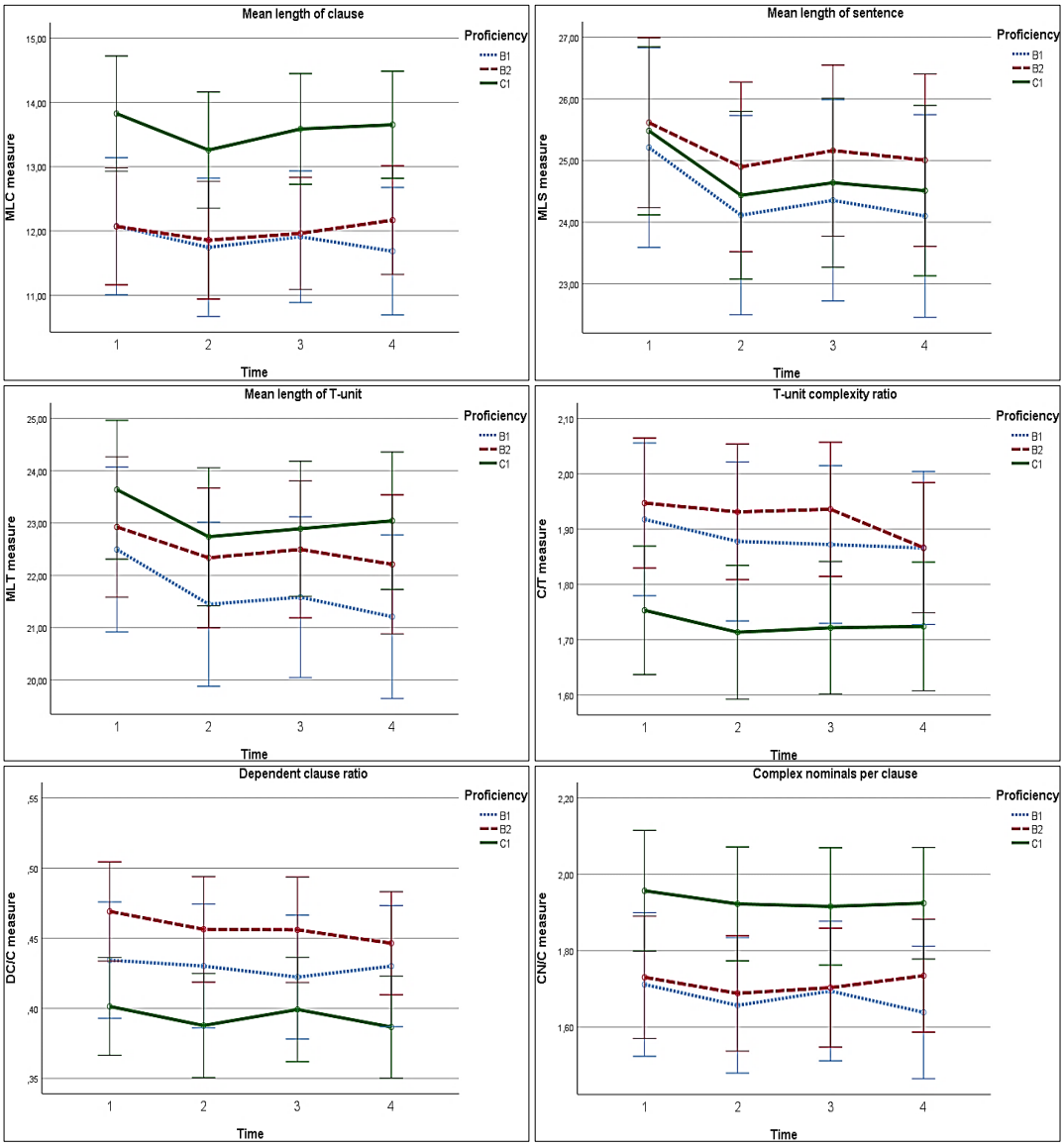
	Measure	III sum of square	*df*	*F*	*Sig.*	Partial Eta Squared
Syntactic complexity	MLC	276.206	2	4.648	.012	.080
	MLS	36.288	2	.243	.785	.005
	MLT	132.228	2	.971	.382	.018
	C/T	3.297	2	3.054	.051	.054
	DC/C	.326	2	3.129	.048	.055
	CN/C	5.647	2	3.041	.052	.054
	CN/T	2.635	2	.600	.551	.011
Readability and language issues	FKGL	115.852	2	4.039	.020	.070
	FRE	5268.261	2	6.816	.002	.113
	LI	6.038	2	.148	.863	.003
Lexical diversity	Tokens	5023.284	2	.540	.584	.010
	Types	4928.835	2	1.708	.186	.031
	TTR	.038	2	.553	.577	.010

Results present significant differences in the syntactic complexity measures of MLC $F_{(2)} = 4.648$, $p = .012$, $\eta_p^2 = .080$. C1-level participants achieved higher mean clause length (M = 13.58, SE = .426) compared to B1- (M = 11.85, SE = .506, $p = 0.31$ [IC 95% 12.737, 14.425]) and B2-level students (M = 12.01, SE = .431, $p = 0.33$ [IC 95% 11.161, 12.870]). Differences in DC/C were found significant in the comparison between B2 (M = .429, SE = .021) and C1 proficiency levels (M = .394, SE = .018, $p = 0.42$ [IC 95% .387, .471]). The dependent variables of *FKGL* ($F_{(2)} = 4.039$, $p = .020$, $\eta_p^2 = .070$) and FRE ($F_{(2)} = 6.816$, $p = .002$, $\eta_p^2 = .113$) significantly differed between B1 and C1 proficiency levels. As for the

Print ISSN: 1578-7044; Online ISSN: 1989-6131

FKGL, C1-level students (M = 13.76, SE = .296) attained higher mean scores than B1-level students (M = 12.47, SE = .352, p = 0.18 [IC 95% 11.775, 13.169]). Regarding FRE measures, the group with C1-level proficiency reached lower mean values (M = 37.86, SE = 1.535) than B1-level participants (M = 46.56, SE = 1.825, p = 0.001 [IC 95% 42.948, 50.185]).

The study of proficiency effects on the dependent variables was further explored, since this reveals variations in participants' revisions progress from their self-revised initial draft (T1) to the final abstract version (T4). Post-hoc Tukey tests were carried out to examine the impact of proficiency on students' outcomes, i.e., *syntactic complexity*, *readability* and *language issues*, and *lexical diversity* measures. Results revealed no significant differences (p > .05) between the three proficiency levels at the different revision stages (T1–T4) for all measures. Estimated marginal means for the dependent variables under the time*proficiency model are presented below (Figures 1–3):



Print ISSN: 1578-7044; Online ISSN: 1989-6131

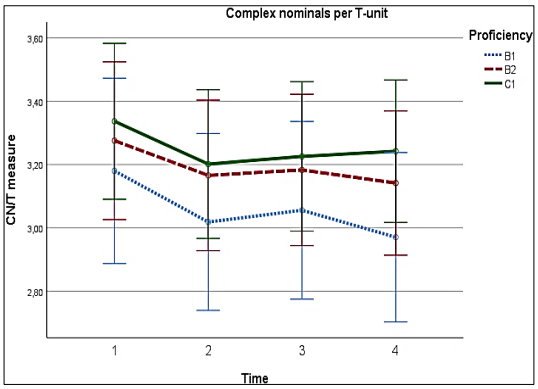
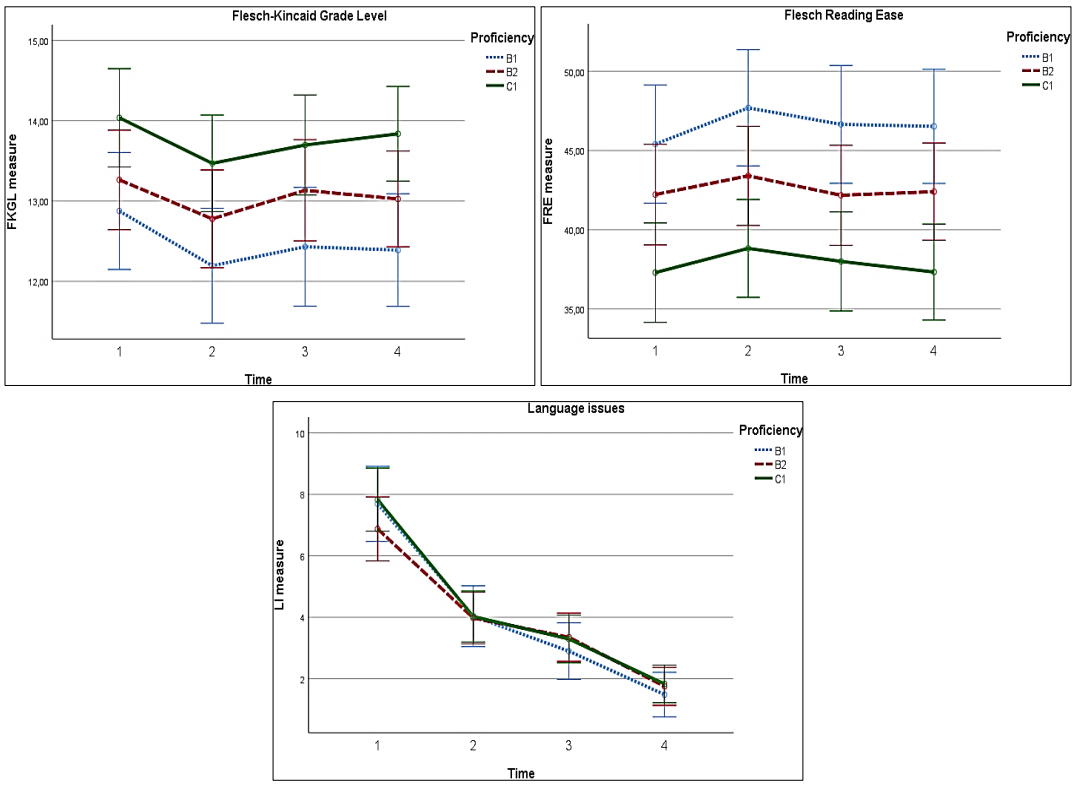


Figure 1. Syntax complexity at different levels of proficiency (B1, B2, C1) and stages of writing revisions (T1–T4).

As shown in Figure 1, the results for *syntactic complexity*, variables revealed an overall descending tendency (T1–T4) regardless of participants' proficiency. The global measures of mean length of clause (MLC), mean length of sentence (MLS), and mean length of T-unit (MLT) were reduced. Likewise, subordination measures i.e., clauses per T-unit (C/T) and the number of dependent clauses per clause (DC/C) presented a decreasing tendency. C1-level participants' results were lower for these two variables, compared with B1- and B2-level students. The measures accounting for complex nominals per clause (CN/C) and per T-unit (CN/T) gradually reduced as well (T1–T4), except for B2-level participants' CN/C, which measured a modest increase at T4.



Print ISSN: 1578-7044; Online ISSN: 1989-6131

Figure 2. Readability and language issues at different levels of proficiency (B1, B2, C1) and stages of writing revisions (T1–T4).

The *readability* measures of FKGL and FRE seem to align well with participants' proficiency levels (Figure 2). The results for FKGL slightly decreased (T1–T2) for all proficiency levels, therefore presenting more complex texts. Nonetheless, these values increased from T2 to T4 but did not return to the initial scores. Similarly, the FRE measures showed an initial tendency to increase (T1–T2) but gradually reduced from the T2 to T4 revision stages. As regards the number of *language issues* (LI), there was a steep downward trend (T1–T2) in the number of language issues (LI) that became consistent (T2–T4) at the different proficiency levels; still, some language issues remained, i.e., punctuation, verb forms, and subordinated clauses.

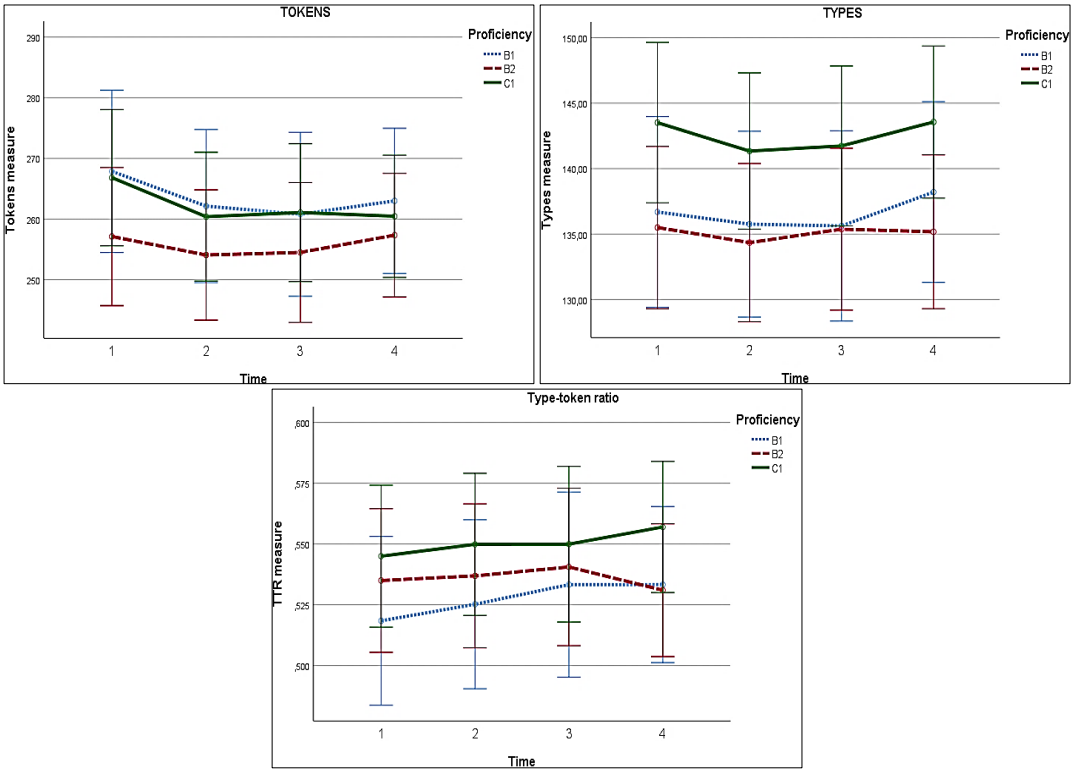


Figure 3. Lexical diversity at different levels of proficiency (B1, B2, C1) and stages of writing revisions (T1–T4).

Lexical diversity measures (Figure 3) depicted an overall initial decrease in the number of tokens (T1–T2), which stabilised (T2–T3) for a posterior marginal increase at B1 and B2 levels. Such stabilisation was less consistent at the C1 proficiency level (T3–T4). The mean number of types initially decreased regardless of participants' proficiency levels (T1–T2). There was a posterior gradual increase at B1 and C1 proficiency levels (T2–T4), which did not apply to B2-level participants (T3–T4). The TTR mean score values slightly increased

Print ISSN: 1578-7044; Online ISSN: 1989-6131

(T1–T3) at B1, B2, and C1 proficiency levels. This tendency remained consistent at C1 proficiency levels (T3–T4) but the B2 TTR ratio levelled off at the T4 revision stage while B1-level students' results marginally stabilized (T3–T4).

5. DISCUSSION

This study aimed at examining the effect of self- and AWE-mediated revision processes on EFL students' writing progress (RQ1). Participants' language proficiency was studied as a mediating factor between AWE feedback provision and EFL students' written outcomes (RQ2). The findings from the first research question were mostly contradictory. Results of the *syntactic complexity* dependent variables confirm an overall decrease over revisions. Thus, there has been no improvement in participants' syntactic complexity (T1–T4) as noted in Huang and Renandya's study (2020). In contrast, Li et al. (2020) report gains in syntactic complexity after exposing college students to AWE feedback. Nevertheless, the minimal mean score differences do not imply a loss in syntactic complexity, as reported in Xu and Zhang (2022) and Thi and Nikolov's (2023) studies.

As regards participants' abstract *readability*, there was marginal improvement in the FKGL variable over revisions. Nonetheless, non-significant differences were identified concerning abstracts' FRE over time. The combination of AWE feedback provision and students' self-revision was found to be statistically significant ($p < .001$, $\eta_p^2 = .694$) in language issues reduction. This decrease in LI and the minor differences in FRE and FKGL mean scores (T1–T4) indicate more accurate writing. These findings confirm students' best efforts to improve writing accuracy, hindering syntactic complexity progress (Thi & Nikolov, 2023). Some of the arguments supporting this view include that ESL–EFL students' best attempts are to avoid making mistakes as a major matter of concern (Truscott, 2007), and their subsequent use of non-complex sentences is to achieve accuracy (Li et al., 2015; Liao, 2016). As for the analysis of *lexical diversity*, the number of tokens and types decreased from T1 to T4. The TTR mean scores, regarded as an indicator of academic writing (Nasseri & Thompson, 2021), revealed a statistically significant increase (T1–T4) ($p < .05$). However, the effect sizes were small–medium ($\eta_p^2 \leq .067$). As indicated by Xu and Zhang (2022), these results point towards a moderate increase in lexical diversity due to participants' access to synonyms as provided in the AWE applications.

For the second research question, significant differences at different proficiency levels were only identified in the mean length of clause (MLC) and the dependent clause ratio (DC/C). C1-level participants produced longer clauses than B1- and B2-level proficient participants with a medium effect size ($\eta_p^2 = 0.80$). B2-level students' dependent clauses ratio (DC/C) was significantly higher than that of students with C1-level proficiency. This is in

Print ISSN: 1578-7044; Online ISSN: 1989-6131

agreement with Casal and Lee's (2019) perspective on complex nominals evidencing complexity at phrase level, which favours the use of less subordinated structures as suggested in Norris and Ortega's (2009) proposal for syntactic development. Concerning *readability* variables, FKGL and FRE results pointed towards increased text complexity at higher proficiency levels; yet the remaining dependent variables did not present significant variability among groups.

The study of proficiency effects over revisions (i.e., T1–T4 revisions) unveiled no significant differences that point to substantial writing improvement. Hence, as previously put forward by Saricaoglu (2019) and Wilson and Roscoe (2020), little or no writing improvement was achieved at the different proficiency levels from T1 to T4. These findings are partly in disagreement with published studies conducive to EFL students' proficiency as determining AWE feedback engagement and exploitation (Koltovskaia, 2020; Zheng & Yu, 2018). On the contrary, our results are consistent with previous research on the little evidence that writing quality progresses when overcoming multiple AWE-mediated revisions (Huang & Renandya, 2020; Stevenson & Phakiti, 2014), and with those of Li et al. (2015), which found that multiple revisions enhanced EFL students' writing accuracy.

While not neglecting L2 students' proficiency as mediating the understanding of the AWE feedback provided (Link et al., 2020; Ranalli, 2018; Thi & Nikolov, 2023), our results further support the potential of AWE tools to reduce language issues (LI) (Lee, 2020; Ranalli et al., 2017; Wang et al., 2013). Several factors have influenced participants' abstract writing and revision practices. First, EFL students' unfamiliarity with the academic writing register has constrained their written outcomes as suggested by Spring et al. (2010), Davidson (2019), and Guo et al. (2021). In this light, Thi and Nikolov (2023) underline the effects of different text typology demands (e.g., communicative intention and language requirements) as challenging students' writing performance. Second, participants' novelty as AWE tools users has hindered feedback engagement and uptake, therefore restricting their corrections to LL issues. This concurs well with the assumptions of Jinxing and Razalli (2020), Ranalli (2018), Huang and Renandya (2020), and Liao (2016). Last, behavioural, cognitive, and affective factors could have influenced students' proper feedback engagement and uptake (Koltovskaia, 2020; Zhang, 2017; Zheng & Yu, 2018).

As for the limitations of the study, the lack of a control group and the number of participants at different proficiency levels could have affected the validity of results to a certain degree. Additionally, the instructional treatment design has lacked time devoted to more extensive instruction in academic writing, revision strategies, and AWE usage, which has determined participants' writing outcomes, as suggested in McCarthy et al. (2022). Likewise, the absence of fit-for-proficiency feedback provision on the part of the language instructor has hindered students' writing progress.

Print ISSN: 1578-7044; Online ISSN: 1989-6131

These findings have implications for EFL language instructors, learners, and software developers. Providing students with additional instruction in the use of AWE tools and strategies for text revision in line with academic writing conventions can positively impact their writing skills and revision practices. This stems from the fact that EFL students require time to understand, interpret, and internalize the academic register specificities along with the AWE feedback provided (Liao, 2016). Furthermore, language instructors' feedback provision at certain revision stages and proficiency levels may enhance students' writing progress. As noted by Zhang and Zhang (2022), the combination of the available feedback sources (i.e., AWE tools and language instructors) caters to different students' needs from LL to HL writing aspects. Lastly, based on the premise that AWE feedback is AI-determined and still lacks pedagogical adaptability (Bitchener & Storch, 2016), software developers need to diversify feedback provision considering writers' proficiency. This should progress from LL aspects to more meaning-focused writing components, enabling EFL students' self-monitor skills development as a step-by-step process, gradually and systematically (Xu & Zhang, 2022).

6. CONCLUSIONS

The evidence from this study supports the idea that instructing EFL learners on AWE tools' usability can positively affect their writing revision practices and outcomes to a certain extent. This was previously indicated by Chapelle et al. (2015), Cotos (2014), and Li et al. (2017), and substantiated by Lee (2020), Lee et al. (2013) and Wang et al.'s (2013) studies. Even though the instructional treatment devised ensures explicit instruction and sustained revision practices (Guo et al., 2021; Strobl et al., 2019), the time devoted to the use of the academic writing register and AWE tools was limited. Moreover, the non-interventionist role of the language instructor as a feedback source has hindered students' progression. Future research should further examine the impact of formative and corrective feedback provided by language instructors, in addition to that of automated writing corrective feedback (AWCF). This is grounded in previous findings in the literature that call for i) the beneficial role of integrating ESL–EFL learners' instruction in using AWE tools –that is, raising awareness and understanding on the uses and limitations of the corrective and diagnostic feedback provided through instructional practices (Chapelle et al., 2015; Cotos, 2014; Rodríguez-Peñarroja, 2022); ii) the crucial importance of undertaking multiple revisions (Barkaoui, 2016; Chambers, 2011; Chapelle & Voss, 2017; Chen & Zhang, 2019; Hayes, 2012; Li et al., 2015; Warschauer & Ware, 2006), which may increase participants' self-efficacy in revision practices; and iii) participants' improved autonomy (Dikli, 2006; Li et al., 2015; Liao, 2016; Wang et al., 2013) and capacity to diagnose and resolve self- and AWE-flagged language misuse (Cotos, 2014; Stevenson & Phakiti, 2014).

Print ISSN: 1578-7044; Online ISSN: 1989-6131

The above-mentioned factors also should encourage writing instructors to give thoughtful consideration of how to best include AWE for formative purposes in ESL–EFL students' writing practices. This needs to consider participants' proficiency, the written genres to be mastered, and the social and cultural specificities of diverse teaching and learning contexts. Needless to say, AWE tools should be seen as a replacement for neither human instruction nor feedback provision (Cotos, 2014; Rodríguez-Peñarroja, 2022), but as supplementary systems that provide corrective and diagnostic feedback. The extensive up-to-date literature on AWE tools invaluably enriches this state-of-the-art topic and may serve as a continuous spur to future research on the use of technology for writing purposes in the EFL and ESL teaching and learning contexts.

REFERENCES

Alameddine, M. M., & Mirza, H. S. (2016). Teaching academic writing for advanced level grade 10 English. *Procedia-Social and Behavioral Sciences, 232*, 209–216. https://doi.org/10.1016/j.sbspro.2016.10.048

Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *Modern Language Journal, 100*(1), 320–340. https://doi.org/10.1111/modl.12316

Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English.* Harlow, Essex: Pearson Education.

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development* (Vol. 96). Multilingual Matters.

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing, 44*, 51–62. https://doi.org/10.1016/j.jslw.2019.03.005

Chambers, L. (2011). Composition and revision in computer-based written assessment. *Research Notes, 43*, 25–32.

Chapelle, C. A., & Voss, E. (2017). Utilizing Technology in Language Assessment. In: Shohamy, E., Or, I., May, S. (Eds.), *Language Testing and Assessment. Encyclopedia of Language and Education*. Springer, Cham. https://doi.org/10.1007/978-3-319-02261-1_10

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing, 32*(3), 385–405. https://doi.org/10.1177/0265532214565386

Chen, C., & Cheng, W. (2008). Beyond the design of automated writing evaluation: pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology, 12*(2), 94–112. http://dx.doi.org/10125/44145

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Chen, J., & Zhang, L. J. (2019). Assessing student-writers' self-efficacy beliefs about text revision in EFL writing. *Assessing Writing, 40*(2), 27–41. https://doi.org/10.1016/j.asw.2019.03.002

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

Council of Europe (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume,* Council of Europe Publishing, Strasbourg. Retrieved January 3, 2024, from http://www.coe.int/lang-cefr

Corazza, A., Maggio, V., & Scanniello, G. (2018). Coherence of comments and method implementations: A dataset and an empirical investigation. *Software Quality Journal, 26*(2), 751–777. https://doi.org/10.1007/s11219-016-9347-1

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. Palgrave Macmillan.

Cotos, E. (2023). Automated Feedback on Writing. In O. Kruse, C. Rapp, C. Anson, K. Benetos, E. Cotos, A. Devitt, A. Shibani (Eds.), *Digital Writing Technologies in Higher Education* (pp. 347–364). Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_22

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing, 26,* 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Davidson, P. (2019). What is academic writing? In W. Zoghbor, S. Al Alami, & T. Alexiou, (Eds.), *Proceedings of the 1st applied linguistics and language teaching and learning in a globalized world (ALLT) Conference* (pp. 12–20). Zayed University Press.

Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment, 5*(1), 1–35. Retrieved December 8, 2023 from https://ejournals.bc.edu/index.php/jtla/article/view/1640

Dizon, G., & Gayed, J. M. (2021). Examining the impact of Grammarly on the quality of mobile L2 writing. *The JALT CALL Journal*, (*17*)2, 74–92. https://doi.org/10.29140/jaltcall.v17n2.336

Eckstein, G., & Bell, L. (2023). Dynamic written corrective feedback in first-year composition: Accuracy and lexical and syntactic complexity. *RELC Journal, 54*(3), 630–647. https://doi.org/10.1177/00336882211061624

Eleyan, D., Othman, A., & Eleyan, A. (2020). Enhancing software comments readability using Flesch reading ease score. *Information, 11*(9), 430. https://doi.org/10.3390/info11090430

Flesch, R. (1979). *How to write plain English: A book for lawyers and consumers*. Harper and Row.

Goldin, I., Narciss, S., Foltz, P., & Bauer, M. (2017). New directions in formative feedback in interactive learning environments. *International Journal of Artificial Intelligence in Education, 27*(3), 385−392. https://doi.org/10.1007/s40593-016-0135-7

Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia-Social and Behavioral Sciences, 198,* 546–556. https://doi.org/10.1016/j.sbspro.2015.07.477

Guo, Q., Feng, R., & Hua, Y. (2021). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 1–20. https://doi.org/10.1080/09588221.2021.1879161

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Herrington, A., & Moran, C. (2012). Writing to a machine is not writing at all. In N. Elliot & L. Perelman (Eds.), *Writing in the 21st Century: Essays in honor of Edward M. White* (pp. 219–232). New York: Hampton Press.

Hockly, N. (2019). Automated writing evaluation. *ELT Journal*, *73*(1), 82−88. https://doi.org/10.1093/elt/ccy044

Huang, S., & Renandya, W. A. (2020). Exploring the integration of automated feedback among lower-proficiency EFL learners. *Innovation in Language Learning and Teaching*, *14*(1), 15−26. https://doi.org/10.1080/17501229.2018.1471083

Jingxin, G., & Razali, A. B. (2020). Tapping the potential of Pigai automated writing evaluation (AWE) program to give feedback on EFL writing. *Universal Journal of Educational Research*, *8*(12B), 8334–8343. https://doi.org/10.13189/ujer.2020.082638

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, *53*, 61–79.

Kaharuddin, A. (2020). Contributions of technology, culture, and attitude to English learning motivation during COVID-19 outbreaks. *Systematic Reviews in Pharmacy*, *11*(11), 76–84.

Kobayashi, H., & Rinnert, C. (2001). Factors relating to EFL writers' discourse level revision skills. *International Journal of English Studies*, *1*(2), 71–101. Retrieved from https://revistas.um.es/ijes/article/view/48111

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*, 100450. https://doi.org/10.1016/j.asw.2020.100450

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Doctoral Dissertation, Georgia State University]. ScholarWorks.

Lee, C., Cheung, W. K. W., Wong, K. C. K., & Lee, F. S. L. (2013). Immediate web-based essay critiquing system feedback and teacher follow-up feedback on young second language learners' writings: An experimental study in a Hong Kong secondary school. *Computer Assisted Language Learning*, *26*(1), 39–60. https://doi.org/10.1080/09588221.2011.630672

Lee, Y-J. (2020). The long-term effect of automated writing evaluation feedback on writing development. *English Teaching*, *75*(1), 67–92. https://doi.org/10.15858/engtea.75.1.202003.67

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*, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Li, W., Lu, Z., & Liu, Q. (2020). Syntactic complexity development in college students' essay writing based on AWE. *CALL for widening participation: Short papers from EUROCALL*, 190–194. https://doi.org/10.14705/rpnet.2020.48.1187

Li, Z. (2021). Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence. *System*, *99*, 102505. https://doi.org/10.1016/j.system.2021.102505

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Li, Z., Dursun, A., & Hegelheimer, V. (2017). Technology and L2 writing. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 77–92). John Wiley & Sons.

Liao, H. C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System, 62*, 77−92. https://doi.org/10.1016/j.system.2016.02.007

Lillis, T., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication, 23*(1), 3–35. https://doi.org/10.1177/0741088305283754

Link, S., Mehrzad, M., & Rahimi, M. (2020). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, *33*, 1–30. https://doi.org/10.1080/09588221.2020.1743323

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, *52*, 100608. https://doi.org/10.1016/j.asw.2022.100608

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Muñoz-Luna R. (2015). Main ingredients for success in L2 academic writing: Outlining, drafting and proofreading. *PLoS ONE, 10*(6), 1–15. https://doi.org/10.1371/journal.pone.0128309

Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, *47*, 100511. https://doi.org/10.1016/j.asw.2020.100511

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Nova, M. (2018). Utilizing Grammarly in evaluating academic writing: A narrative research on EFL students' experience. *Premise: Journal of English Education and Applied Linguistics 7*(1), 80–96. http://dx.doi.org/10.24127/pj.v7i1.1332

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning*, *31*(7), 653−674. https://doi.org/10.1080/09588221.2018.1428994

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*(1), 8–25. https://doi.org/10.1080/01443410.2015.1136407

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational research review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rodríguez-Peñarroja, M. (2022). Automated writing evaluation: Present challenges and future directions. In N. Martínez León, I. Mosquera Gende, & A. J. Moreno Guerrero (Eds.), *Tic Docentes* (pp. 357–370). Thomson Reuters Aranzadi.

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Saricaoglu, A. (2019). The impact of automated feedback on L2 learners' written causal explanations. *ReCALL*, *31*(2), 189–203. https://doi.org/10.1017/S095834401800006X

Sengupta, S. (2000). An investigation into the effects of revision strategy instruction on L2 secondary school learners. *System*, *28*(2), 97–113. https://doi.org/10.1016/S0346-251X(99)00063-9

Soleimani, H., & Moqimi, J. (2023). Teaching ESL/EFL Writing Skill Website Evaluation: The Purdue Online Writing Lab and Pro Writing Aid. *Interdisciplinary Studies in English Language Teaching*, *2*(1), 127–145. https://doi.org/10.22080/iselt.2023.25121.1049

Spring, J., Davidson, P., & Palmer, D. (2010). Why we should not be teaching academic writing through the rhetorical pattern approach. In D. Anderson & C. Coombe (Eds.), *Cultivating real writers: Emerging theory and practice for adult Arab learners* (pp. 53–64). HCT Press.

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, *19*, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Stevenson, M., Schoonen, R., & de Glopper, K. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*, *15*(3), 201–233. https://doi.org/10.1016/j.jslw.2006.06.002

Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, *131*, 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

Thi, N. K., & Nikolov, M. (2023). Effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(1), 6. https://doi.org/10.1186/s40862-022-00182-1

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of second language writing*, *16*(4), 255–272. https://doi.org/10.1016/j.jslw.2007.06.003

University of Cambridge Local Examinations Syndicate (UCLES) (2001). *Quick placement test*. Oxford University Press.

Wallwork, A. (2011). *English for writing research papers*. Springer.

Wang, Y. J., Shang, H. F., & Briody, P. (2013). Exploring the impact of using automated writing evaluation in English as a foreign language university students' writing. *Computer Assisted Language Learning*, *26*(3), 234–257. https://doi.org/10.1080/09588221.2012.655300

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, *10*(2), 157–180. https://doi.org/10.1191/1362168806lr190oa

Waycott, J., Bennett, S., Kennedy, G., Dalgarno, B., & Gray, K. (2010). Digital divides? Student and staff perceptions of information and communication technologies. *Computers & Education*, *54*(4), 1202–1211. https://doi.org/10.1016/j.compedu.2009.11.006

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87−125. https://doi.org/10.1177/0735633119830764

Xu, J., & Zhang, S. (2022). Understanding AWE feedback and English writing of learners with different proficiency levels in an EFL classroom: A sociocultural perspective. *The Asia-Pacific Education Researcher*, *31*(4), 357–367. https://doi.org/10.1007/s40299-021-00577-7

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Zhang, J., & Zhang, L. J. (2022). The effect of feedback on metacognitive strategy use in EFL writing. *Computer Assisted Language Learning*, *37*(5–6), 1–26. https://doi.org/10.1080/09588221.2022.2069822

Zhang, Z. V. (2017). Student engagement with computer-generated feedback: A case study. *ELT Journal*, *71*(3), 317–328. https://doi.org/10.1093/elt/ccw089

Zhang, Z. V. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*, 100439. https://doi.org/10.1016/j.asw.2019.100439

Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90–102. https://doi.org/10.1016/j.asw.2018.02.004

Zheng, Y. & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing*, *37*, 13−24. https://doi.org/10.1016/j.asw.2018.03.001

Print ISSN: 1578-7044; Online ISSN: 1989-6131