Advanced Journal of Interdisciplinary Studies
Volume 10, Issue 1, 2026 | Pages 3099–3116
Journal homepage: www.ajids.dmu.edu.et

Research Article

Effects of Reflective Writing Tasks on EFL Learners' Engagement in Writing

Zewudu Paulos Mada[1], Sualih Mussa Akalu[1], Getinet Bilew Fenta[1]

[1]*Department of English Language and Literature, Debre Markos University,P.O. Box 269, Debre Markos, Amhara, Ethiopia*

ARTICLE INFO

Received: 30 March 2026
Revised: 2 October 2026
Accepted: 6 October 2026
Available online: 8 October 2026

KEYWORDS

Reflective writing tasks;
Students' engagement;
Behavioral engagement;
Cognitive engagement;
Emotional engagement;
Social engagement;
Interrupted time-series design

ABSTRACT

Student engagement is widely recognized as central to developing writing skills in a target language, yet little is known about how Reflective Writing Tasks (RWTs) shape engagement among learners of English as a Foreign Language. This study examined the effect of an eight-week RWT intervention on four dimensions of EFL writing engagement (behavioral, cognitive, emotional, and social) and explored students' lived experience of the intervention. A convergent mixed-methods design was used. The quantitative strand was a one-group Interrupted Time-Series Design with six measurement points (three before and three after the intervention), in which 30 third-year students at Jinka University completed a 20-item engagement questionnaire based on Fredricks et al. (2004). The data were analyzed using a one-way repeated-measures ANOVA with Bonferroni-adjusted pairwise comparisons and polynomial trend analysis. The qualitative strand comprised semi-structured interviews with eight purposively selected participants, analyzed through thematic analysis (Braun & Clarke, 2021). Time had a large effect on all four engagement dimensions (partial η^2 = .775–.837, $p < .001$): scores were stable across the baseline, rose sharply between the third pre-intervention and first post-intervention measurements, and continued to increase thereafter. The interviews helped to explain this pattern, as participants reported greater effort and persistence, more strategic learning, reduced writing anxiety, stronger peer cooperation, and more productive tutoring interactions with the instructor. Because the design has no control group and the sample is small and drawn from a single institution, the findings are preliminary and do not establish causation; they do, however, suggest that RWTs are a promising approach to enhancing EFL writing engagement that merits testing in controlled studies.

1. Introduction

1.1. Background of the study

One of the major challenges facing the higher education system in Ethiopia is the need to improve advanced level writing skills. Taking into consideration the importance of English language throughout the university curriculum (Ministry of Education, 2023), English writing is not only a mode of communication but also a means of showing knowledge and creating an academic identity. However, there are several challenges in developing such writing

Corresponding author: zewudupaul4@gmail.com *(Mada Z.P.)*

DOI: http://doi.org/10.20372/ajids.2026.3099

skills due to a range of issues including product-oriented teaching approaches and lack of evidence-based frameworks.

At the same time, student engagement has been recognized worldwide as a strong predictor of academic achievement, persistence, and quality of learning (Fredricks et al., 2004; Hınız & Çelik, 2025). The construct, which covers involvement of the individual at the behavioral, cognitive, emotional, and social levels in academic activities, has been extensively studied in the field of first- and second-language learning (Al-Hoorie, 2014; Hegarty et al., 2026). Despite that, there seems to be a clear deficiency in the knowledge about the association between student engagement and writing instruction, with the former being recognized but not much research being conducted in this area.

Reflective Writing Tasks (RWTs) can be considered a theoretically sound method of increasing engagement. Using Dewey's (1933) theory of experiential learning and Schön's (1983, 1987) concepts of reflection in action and reflection on action, RWTs allow learners to take control over their learning process by reviewing, evaluating, and revising the results and the process of writing. This method is theoretically based on the idea that reflection on the decision-making process during the writing, along with self-monitoring and self-explanation, allows activating effort, cognition, emotion, and socialization in regard to the task of writing.

Despite its relevance, empirical support for the effectiveness of RWTs within the context of EFL writing engagement is still relatively lacking, especially when it comes to sub-Saharan African countries such as Ethiopia. Much of the previous research has been concerned with the effects of such interventions on writing quality, which took place in different higher education settings, while the experimental designs used were not effective in controlling for history, maturation, or even testing effects (Kumari & Prakasha, 2025; Rizqiani & Astuti, 2025).

1.2. Problem Statement

The current literature shows that the paradigm currently in place in writing instruction at Ethiopian institutions of higher education is primarily product-based, lacking adequate focus on learner autonomy, metacognition, and engagement (Woldemariam, 2018). Even though learners typically act as receivers of feedback on the completed written pieces, there has been no effort put into creating scaffolding which would enable the learners to engage in reflective thinking and self-monitoring, a process that can be expected to lead to improvement of their writing skills. Deficiencies in engagement among students have been noted in many aspects, including behavioral (superficial compliance), cognitive (reproductive learning strategies), emotional (writing anxiety), and social (poor peer interaction).

There is a well-developed theoretical foundation for considering reflective writing as a tool which may enhance engagement. However, what is lacking in the literature is a body of research that would explore this problem with the inclusion of both trend analysis and the point of view of the learners. Without that information, any discussion about innovative teaching methods is bound to remain abstract. The current study seeks to provide an example of empirical research of that kind.

1.3. Research Objectives

1.3.1. General Objective

To assess the effect of Reflective Writing Tasks (RWTs) on the multidimensional engagement of EFL learners in an English writing course in one of higher education at Jinka University.

1.3.2. Specific Objectives

1. To assess the effect of the RWTs intervention on EFL learners' multidimensional engagement before and after the intervention implementation.

2. To investigate whether changes in the learners' engagement after the implementation of the RWTs intervention are sustainable and find the engagement dimension that benefits most from the implementation.

3. To discover the students' experiences and perceptions of the intervention and find out whether they relate to the quantitative results of the intervention regarding engagement.

1.4. Significance of the Study

There are three key goals in the current study. First, based on the literature review, the paper discusses one of the earliest studies of the effects of the writing task intervention (RWTs) on the engagement of multivariate English as a Foreign Language (EFL) learners as the major variable in a higher educational institution using longitudinal design with six measures and interviews. Second, this paper uses polynomial contrasts in the repeated-measures analysis to outline the course of engagement in more detail compared to previous works about EFL writing intervention. Third, the paper considers engagement as an operationalizable variable in teaching.

2. Literature Review

2.1. Theoretical Framework: From Reflection to Engagement

The theoretical foundation of the current study is based on two academic traditions. The first one is the reflective practice tradition. It implies that the reflective practice helps to facilitate learning by reconstructing experience through reflection. This concept was introduced by Dewey (1933) and further developed by Schön (1983, 1987), who defined the concept of reflection-in-action and reflection-on-action as real-time modification of actions and retrospective analysis of actions, correspondingly. In terms of the writing instruction discourse, reflective practice can be considered as a part of writing development rather than an auxiliary tool.

The second tradition, which serves as a theoretical framework in this study, is the multi-dimensional model of student engagement introduced by Fredricks, Blumenfeld, and Paris (2004) and further developed. This framework states that the engagement of students is multidimensional. It involves behavioral, cognitive, emotional, and the recently added social aspects. Behavioral engagement implies effort, participation, and persistence. Cognitive engagement includes self-regulation and deep processing. Emotional engagement stands for affective involvement, interest, and anxiety. Social engagement is related to interactions with peers and tutors.

Considering the characteristics of the reflective processes in writing, the potential parallels between rationalized writing task interventions and multi-dimensional student engagement can be identified. Behaviorally, RWTs require sustained participation in successive cycles of writing and reflection. At the same time, RWTs include planning, monitoring, and self-reflection activities, which can be associated with deep cognitive engagement. Emotionally, reflective cycles can provide opportunities for gradual skills development with reduced anxiety and improved self-efficacy.

2.2. Empirical Evidence on Reflective Writing and Engagement

The existing body of knowledge on RWTs and engagement is rather extensive and growing, but much of it shows some weaknesses in terms of coherence and/or research methodology. First of all, several works have reviewed the empirical literature on engagement in language learning and found out that reflection is positively related to both behavioral and cognitive types of engagement, but noted the lack of longitudinal designs that could allow making causal inferences about the effects under investigation. Second, bibliometric analysis reveals that reflective writing has become one of the most studied interventions aimed at increasing engagement in English language teaching; the dominating design in such research is cross-sectional and based on self-report measures.

A number of empirical studies have concentrated on RWTs specifically. Thus, Teng and Huang (2026) found that EFL learners receiving reflective feedback were more behaviorally and emotionally engaged than those getting only traditional corrections. Bui and Nguyen (2026) reported that reflecting writing in dialogues increased social engagement in EFL learners during grammar instruction. Liang and Wu (2025) using ecological momentary assessment showed that academic self-efficacy of students developed in accordance with the pattern of reflection. Finally, Ceylan and Korkmaz (2026) revealed the positive correlation between engagement and academic achievement in EFL environments; thus, it can be stated that engagement is a very important result to study. As far as empirical studies from Ethiopia or sub-Saharan Africa are concerned, there is not much literature on the topic.

2.3. The Interrupted Time-Series Design (ITSD) in Educational Research

The Interrupted Time Series Design (ITSD) is a form of quasi-experiment, often used in educational intervention studies where randomization is not possible (Mark & Gamble, 2009; Marsden & Torgerson, 2012). The use of multiple measures taken before implementation of the educational intervention allows the assessment of whether there are any trends pre-intervention, reducing the threats of internal validity such as maturation, testing, regression, and history by means of other plausible explanations for the changes in students' performance after the intervention. Additionally, having a number of post-intervention measures helps to distinguish the effect of novelty from the real effect.

Such a design has been used in studies assessing the effectiveness of drugs and public health policies (Hansen & Rossow, 2007). In the case of its application in educational intervention research, such a design has been characterized as enhancing causal exploratory analysis in classroom settings but lacking the power of inference as randomized trials (Marsden & Torgerson, 2012). At least six measures are required in ITSD: three before and three after the intervention. As mentioned in section 7, lack of a control group does not allow completely rejecting some alternative explanations.

3. Methodology

3.1. Research Paradigm and Design

A convergent mixed methods approach was adopted, whereby the Quantitative Interrupted Time Series Design (ITSD) was combined with qualitative interviews. For the quantitative research method, the emphasis was put on identifying changes in students' level of engagement by using repeated measurements, whereas the qualitative side explored the experience of the participants with the RWT intervention. Data from each modality were analyzed independently before the results could be integrated in order to look for convergence or divergence between them.

The quantitative research methodology involved a one-group ITSD design, which included six measurement points. Baseline data were collected at three different time points: Pre-intervention 1 (week 2), Pre-intervention 2 (week 3), and Pre-intervention 3 (week 4). After an interruption occasion caused by an intervention, data were collected at Post-intervention 1 (week 6), Post-intervention 2 (week 7), and Post-intervention 3 (week 8). The intervention itself lasted eight weeks, and it was preceded by interruption after the third point of baseline measurements.

3.2. Participants and Sampling

In terms of quantitative data, there were 30 third-year undergraduate students enrolled in the course on Advanced Writing Skills within the Department of English Language and Literature at Jinka University, Ethiopia, who comprised the sample for the 2024/2025 academic year. In this connection, purposive sampling was applied for two reasons: (a) third-year students are supposed to have a certain language competence to work with structured reflective activities and experience some developmental changes; and (b) the content of the course includes such aspects as argumentation, coherence, lexical richness, and grammatical accuracy, which can be developed via reflection on the Reflective Writing Task (RWT). In total, all 30 participants took part in measurements six times without losing any participant. It should be mentioned that the sample is a purposely chosen group of people from one university and one course; therefore, the results obtained here cannot be generalized to other situations. As it was already mentioned, the results of the research may be replicated (see Section 7).

As far as qualitative data are concerned, eight participants (P1-P8) were purposively chosen from the initial group of 30 people with maximum variation. To be more exact, three participants with low engagement scores (lowest), three with medium engagement scores (medium), and two with high engagement scores (highest) before the intervention were interviewed. These participants took part in qualitative interviews in accordance with their initial engagement scores (mean scores across dimensions measured at Pre-intervention 3). In general, this type of sampling was used not to reflect the proportional sample, but to explore various experiences of participants. The interviews continued until thematic saturation (see Section 4.6) of major themes was achieved, which happened in seven interviews; the eighth interview was carried out to be sure.

3.3. The Reflective Writing Task (RWT) Intervention Protocol

The Reflective Writing Task (RWT) intervention that was carried out throughout the eight-week intervention study followed the guidelines established in the literature (Pavlovich et al., 2009; Kurniawan, 2024) and was incorporated into the curriculum of the Advanced Writing I course (three sessions of 50 minutes per week), as opposed to being an after-class training exercise. The RWT intervention is comprised of four components that are discussed below.

Weekly cycle format. Throughout the eight-week intervention period, a three-sessions-per-week cycle was consistently used. Session 1 (Monday) entailed the students composing a persuasive essay consisting of 250-300 words on various topics provided in the course syllabus (e.g., "Should university students be obliged to do community service before graduation?"). Session 2 (Wednesday) entailed essay exchange between the rotating peer and the peer-review cycle (discussed further below), as well as completing a structured reflective journal based on the draft essay. Session 3 (Friday) entailed getting written feedback from the instructors regarding the previous week's writing task and engaging in a short one-to-one instructor-mediated feedback discussion (discussed further below) prior to beginning the next writing task.

Reflective journaling. After completing each writing task, students composed a one-page reflective journal using a modified version of the Reflective Cycle by Gibbs (1988) that comprised four questions: (1) Description – "What did you write about and how?"; (2) Evaluation – "What did you like about your essay and what did not work?"; (3) Analysis – "Why were the particular organizational pattern, vocabulary, and grammatical structure chosen?"; (4) Action Planning – "How can you improve your next essay?" Journal entries were written in English, took approximately 15-20 minutes, and were submitted along with the assignment to the instructors.

Peer review cycles. Each writing assignment was assessed by a different rotating peer partner (rotated weekly to ensure maximum collaboration in eight weeks). Reviewers filled in a structured feedback form, which had three parts: strengths (two specific observations); weaknesses (two specific observations related to the rubric for Task Response, Organization, Lexical Resource, and Grammatical Accuracy); and a rationale of 50 words or more explaining the assessment of the paper. Peer-review session took approximately 20 minutes and were carried out in pairs during Session 2.

Rubric-based self-assessment. Before turning each writing assignment to the instructor, students completed self-assessment of their writing assignment drafts according to a simplified four-band version of the IELTS Writing Task 2 rubric. Four bands were considered: task response, coherence and cohesion, lexical resource, and grammatical range and accuracy. Students gave a band score from 4 to 9 and justifications of their scores in one sentence. This part of the protocol served two purposes: making students acquainted with the assessment criteria and bridging novice-expert gap.

Instructor-mediated feedback dialogues took place at intervention processes. After receiving the instructor's feedback in written form consistent with the same four-band rubric within one week since the assignment submission, students had five to seven-minute dialogues with the instructor. Dialogues took place in Session 3. Dates of these sessions were rotated among participants (about 3–4 students per session). Thus, each participant had two to three dialogues during an eight-week intervention period with written feedback between dialogues. Within dialogues, students paraphrased the instructor's comments, chose one specific modification they would make to their assignments and formulated one goal for future writing assignments. In order to avoid any power issues related to instructor-facilitated dialogues (see 3.7), a standard list of prompts was used in all dialogues. Participants were aware that participation and content of the dialogues would not affect their course grades, and a research assistant who did not participate in grading selected randomly the dialogues for observation.

Overall, the protocol included eight writing assignments, eight reflective journals (one per assignment), self-assessment of each assignment, eight peer feedbacks in total (as reviewer and reviewee), written feedback from the instructor and two to three dialogues. Fidelity of implementation was estimated on a weekly basis by reviewing randomly selected 20% of reflective journal entries and peer-feedback forms using a pre-defined checklist.

3.4. Engagement Measurement Instrument

Engagement was measured through a 20-item self-report questionnaire modified from Fredricks et al. (2004), and supplemented with an additional five items in a Social Engagement subscale, thus generating four subscales of five

items each (Behavioral Engagement – effort, participation, and persistence; Cognitive Engagement – deep processing, self-regulation, and metacognition; Emotional Engagement – interest, enthusiasm, and writing anxiety, reverse-scored; and Social Engagement – quality of peer and instructor interactions). The items were scored on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree).

This measuring tool exhibited satisfactory psychometric characteristics among the participants in this study. Its internal consistency (Cronbach's alpha) was .792 for Social Engagement (pre-intervention) and .881 for Social Engagement (post-intervention), both above the conventional threshold of .70 required for use in research (see Table 1). The test-retest reliability for all four subscales across the three pre-intervention occasions was very high, with inter-occasion correlations between .79 and .89 ($p < .001$ in all cases), thus attesting to the stability of the baseline measurement (see Table 2). The normality of all engagement dimensions at Pre-intervention 1 was supported by the results of the Shapiro-Wilk test (all $p > .05$), and the sphericity assumption was also met for all dimensions (Mauchly's W from .700 to .893, all $p > .05$), thus eliminating the need for the Greenhouse-Geisser correction for any engagement analysis.

3.5. Qualitative Data Collection

After the intervention had been completed over the course of eight weeks, followed by the third and final (Post-intervention 3) survey, individual semi-structured interviews with a length of 25–35 minutes each were conducted with the eight purposively selected participants (P1-P8) mentioned in Section 3.2. The interviews were carried out either in Amharic or English, recorded with consent of participants and transcribed verbatim; portions in Amharic were translated into English by the first author and double-checked by another bilingual person, not involved in the study, to assure accuracy of translation. In order to avoid possible participants' pressure to respond in the required way, interviews were conducted by a research assistant who was not responsible for assessing the course grade, at the venue other than the one where the course instructor works (see Section 3.7).

Open-ended questions on four engagement dimensions were used as the basis of the interviews protocol. Example questions include: "Can you describe how your approach towards performing writing assignments changed, if it did, during the period of eight weeks?" (behavioral dimension); "Can you give an account of how you used to plan and revise a piece of writing during the last weeks of the course compared with the beginning?" (cognitive dimension); "How did you feel about being provided with feedback from your peers and the instructor at the beginning of the course, and how did your feelings change, if they did?" (emotional dimension); and "How did you find the peer-review and feedback-dialogue sessions?" (social dimension).

3.6. Data Analysis

Quantitative data were analyzed using SPSS Version 27. One-way repeated-measures ANOVA was conducted for each of the four engagement dimensions using TIME as a within-subjects factor (six levels: Pre-intervention 1-3 and Post-intervention 1-3). Partial eta squared (ηp^2) was used as an effect size measure with conventional cut-off points for small ($\eta p^2 = .01$), medium ($\eta p^2 = .06$), and large effect sizes ($\eta p^2 = .14$) based on Cohen (1988). However, effect sizes calculated in studies where participants form a single group without control conditions should be treated cautiously (see Section 5).

Pairwise post hoc comparisons were made with Bonferroni correction to identify the points of significant differences between the six time points. The pattern of significant and non-significant comparisons, stable pre-intervention comparisons, pre-post comparisons, and post-intervention comparisons was compared to the ITSD temporal signature explained in Section 2.3. Polynomial contrasts analysis was conducted for the dimension of Social Engagement (a dimension with the largest effect size) to decompose the trajectory into linear, quadratic, cubic and higher order components.

Interview transcripts were analyzed using reflexive thematic analysis according to six phases developed by Braun and Clarke (2021): (1) familiarization through reading transcripts; (2) initial line-by-line coding; (3) collating codes into candidate themes; (4) revisiting candidate themes against the coded material and full dataset; (5) defining and naming themes; and (6) selecting illustrative extracts. Another researcher (the second author) checked the coded transcripts and candidate themes independently; all discrepancies were discussed until a consensus was reached. Four themes were developed, generally matching the four quantitative dimensions of engagement but being open to the appearance of sub-themes that were not expected on the basis of the quantitative framework.

3.7. Ethical Considerations

This research was approved by the Research Ethics Review Committee of the Department of English Language and Literature, Jinka University (to provide the reference number upon finalization of the research). All 30 students of the Advanced Writing Skills course were informed before the start of the semester about the voluntary character of their participation in the research, including completion of engagement questionnaires at six occasions and, for some of them, participation in interviews at the end of the semester. The students were told that failure to participate would have no impact on their access to RWT activities (included in the course for pedagogical purposes) nor on the assessment of their course performance, and that they could withdraw their data at any moment without any negative consequences. Informed consent was received from all 30 students for the data from the questionnaires and from the eight students interviewed.

Since the first author was the course instructor, there were implemented certain steps in the research design to minimize any possible issues related to power dynamics: (a) administration of questionnaires at all six points of measurement was done by a research assistant, using the sealed anonymous forms marked only by participant codes; (b) interviews were carried out by the research assistant rather than the instructor, in a place other than the class room; (c) students were informed verbally and in writing repeatedly that their responses would remain anonymous with no possibility for the instructor to identify them and that this would have no impact on their grades; and (d) the data collected were immediately anonymized, while the code connecting participant code with identity of the student was known only to the research assistant and destroyed after data collection. Nevertheless, the dual character of the first author's involvement remains a limitation of the design, which is further discussed in Section 7.

3.8. Use of Artificial Intelligence Tools

Generative AI tools, including but not limited to a large language model assistant, were used in the creation of this paper in the following restricted ways: (a) editing the author' rough copy for the sake of grammar and stylistic improvement in order to make it more grammatically correct and clear, as well as better suited for publication in a scholarly journal; (b) formatting the reference list in line with APA 7th Edition rules based on the bibliographic information from the authors; and (c) finding and creating summaries of potential literature for the review section that was later verified by the authors in the original articles. None of the AI tools were used to create research data, conduct statistical analysis, create fake quotes from participants, or generate the actual content of this paper.

4. Results

4.1. Descriptive Statistics for All Engagement Dimensions

Descriptive statistics comprising the mean, standard deviation, and 95% confidence intervals are provided in Table 3 for all four engagement dimensions on all six measures. It should be noted that a pattern is consistent throughout all dimensions and consists of the pre-intervention plateau with moderate engagement and overlapping confidence intervals and subsequent growth at Post-intervention 1 sustained to Post-intervention 3 with a slight additional growth.

Table 1. Cronbach's Alpha Reliability Coefficients for Engagement Dimensions (N = 30)

Dimension	Pretest α	Posttest α	Number of Items	Interpretation
Behavioral Engagement	.808	.818	5	High Reliability
Cognitive Engagement	.870	.848	5	High Reliability
Emotional Engagement	.836	.847	5	High Reliability
Social Engagement	.792	.881	5	High Reliability

From the data collected, it can be stated that there is some clustering around a narrow range from 2.75 (Emotional Engagement, Pre-intervention 1) to 3.10 (Cognitive Engagement, Pre-intervention 2). Thus, the baseline level of engagement on all four dimensions can be considered relatively similar and moderate. At the same time, the pre-intervention standard deviations are relatively low – 0.24 to 0.40. Furthermore, their 95% confidence intervals are overlapping with those of all four dimensions for three pre-intervention measures. The post-intervention means are

estimated at 3.98 (Emotional Engagement, Post-intervention 1) and 4.41 (Behavioral Engagement, Post-intervention 3). Thus, there is a movement from 1.23 to 1.63 scale points. Besides, post-intervention standard deviations are similar to pre-intervention ones – 0.28 to 0.42.

Table 2. Test-Retest Reliability of the Engagement Questionnaire Across Three Pre-Intervention Occasions

Dimension	Pretest 1-2 Correlation (r)	Pretest 1-3 Correlation (r)	Pretest 2-3 Correlation (r)	All p-values	Interpretation
Behavioral Engagement	.86	.83	.88	$p < .001$	Excellent stability
Cognitive Engagement	.84	.81	.87	$p < .001$	Excellent stability
Emotional Engagement	.82	.79	.85	$p < .001$	Good stability
Social Engagement	.85	.82	.89	$p < .001$	Excellent stability

Note. r = Pearson correlation coefficient. All p-values are Bonferroni-corrected.

Table 3. Descriptive Statistics for All Engagement Dimensions Across Six Measurement Occasions

Dimension	**Phase**	**Time Point**	**Mean**	**SD**	**95% CI**
Behavioral Engagement	Pre-Intervention	PRE 1	2.78	0.31	[2.67, 2.90]
		PRE 2	2.87	0.36	[2.73, 3.00]
		PRE 3	2.90	0.40	[2.75, 3.05]
	Post-Intervention	POST 1	4.18	0.35	[4.05, 4.31]
		POST 2	4.13	0.35	[4.00, 4.26]
		POST 3	4.41	0.36	[4.28, 4.54]
Cognitive Engagement	Pre-Intervention	PRE 1	2.91	0.33	[2.79, 3.03]
		PRE 2	3.10	0.33	[2.97, 3.22]
		PRE 3	3.09	0.32	[2.97, 3.21]
	Post-Intervention	POST 1	4.21	0.35	[4.08, 4.34]
		POST 2	4.18	0.28	[4.08, 4.29]
		POST 3	4.34	0.38	[4.20, 4.48]
Emotional Engagement	Pre-Intervention	PRE 1	2.75	0.35	[2.62, 2.88]
		PRE 2	2.82	0.37	[2.69, 2.96]
		PRE 3	2.95	0.27	[2.85, 3.05]
	Post-Intervention	POST 1	3.98	0.42	[3.83, 4.14]
		POST 2	4.07	0.36	[3.94, 4.20]
		POST 3	4.16	0.37	[4.02, 4.29]
Social Engagement	Pre-Intervention	PRE 1	2.84	0.32	[2.72, 2.96]
		PRE 2	2.92	0.36	[2.78, 3.05]
		PRE 3	2.88	0.24	[2.79, 2.97]
	Post-Intervention	POST 1	4.11	0.37	[3.97, 4.25]
		POST 2	4.23	0.32	[4.11, 4.35]
		POST 3	4.29	0.33	[4.16, 4.41]

Note. M = Mean; SD = Standard Deviation; CI = Confidence Interval. N = 30 at all occasions (no attrition). Pre = Pre-intervention; Post = Post-intervention.

4.2. Assumption Testing

Results of Mauchly's Test of Sphericity are presented in Table 4. Sphericity was supported for all four engagement dimensions (all $p > .05$), indicating that within-subjects' variances across the six occasions were sufficiently homogeneous to proceed with standard repeated-measures ANOVA without the Greenhouse-Geisser or Huynh-Feldt corrections. Normality was supported for all dimensions at Pre-intervention 1 via the Shapiro-Wilk test (all $p > .05$): Behavioral Engagement, $W(30) = .951$, $p = .172$; Cognitive Engagement, $W(30) = .962$, $p = .318$; Emotional Engagement, $W(30) = .947$, $p = .134$; Social Engagement, $W(30) = .958$, $p = .267$ (Table 4b).

Table 4. Mauchly's Test of Sphericity for All Engagement Dimensions

Dimension	Mauchly's W	Approx. Chi-Square	df	Sig. (p)	Greenhouse-Geisser ε	Decision
Behavioral Engagement	.778	6.79	14	.943	.893	Sphericity Assumed
Cognitive Engagement	.700	9.66	14	.788	.894	Sphericity Assumed
Emotional Engagement	.661	11.28	14	.665	.876	Sphericity Assumed
Social Engagement	.643	11.96	14	.612	.875	Sphericity Assumed

Note. GG ε = Greenhouse-Geisser epsilon. All $p > .05$, supporting the sphericity assumption; no correction was applied.

Table 4b. Tests of Normality (Shapiro-Wilk) for Engagement Dimensions at Pre-Intervention 1

Dimension	Statistic	Df	Sig. (p)	Decision
Behavioral Engagement	.951	30	.172	Normal
Cognitive Engagement	.962	30	.318	Normal
Emotional Engagement	.947	30	.134	Normal
Social Engagement	.958	30	.267	Normal

4.3. One-Way Repeated-Measures ANOVA: Main Effects

Table 5 presents the within-subjects effects for all four engagement dimensions. All four dimensions showed statistically significant time effects, with effect sizes that are large by Cohen's (1988) conventional criteria for repeated-measures designs. As discussed in Section 5, effect sizes of this magnitude in a single-group design without a control group should be interpreted as indicating a substantial pattern of change over time within this cohort, rather than as a direct estimate of the causal effect of the intervention; they are reported here descriptively, consistent with the ITSD literature (Marsden & Torgerson, 2012).

Table 5. Tests of Within-Subjects Effects for All Engagement Dimensions (One-Way Repeated-Measures ANOVA)

Dimension	Source	SS	Df	MS	F	Sig.(p)	Partial η^2	Power
Behavioral Engagement	TIME	88.55	5	17.71	133.75	<.001	.822	1.000
	Error(TIME)	19.20	145	0.13				
Cognitive Engagement	TIME	67.36	5	13.47	128.68	<.001	.816	1.000
	Error(TIME)	15.18	145	0.11				
Emotional Engagement	TIME	68.74	5	13.75	99.97	<.001	.775	1.000
	Error(TIME)	19.94	145	0.14				
Social Engagement	TIME	80.15	5	16.03	148.80	<.001	.837	1.000
	Error(TIME)	15.62	145	0.11				

Note. SS = Type III Sum of Squares. df = degrees of freedom. All F-values were computed on 5 and 145 degrees of freedom. Sphericity assumed for all dimensions (cf. Table 4).

Table 6. Summary of One-Way Repeated-Measures ANOVA Results with 95% Confidence Intervals for ηp^2

Dimension	Sphericity (ε)	Correction	F (df)	P	Partial η^2	95% CI for η^2	Effect Size Rank
Social Engagement	.875	None (assumed)	F(5, 145) = 148.80	<.001	.837	[.745, .894]	1
Behavioral Engagement	.893	None (assumed)	F(5, 145) = 133.75	<.001	.822	[.726, .882]	2
Cognitive Engagement	.894	None (assumed)	F(5, 145) = 128.68	<.001	.816	[.717, .877]	3
Emotional Engagement	.876	None (assumed)	F(5, 145) = 99.97	<.001	.775	[.668, .842]	4

Note. All F-values are based on df = 5, 145. ηp^2 = partial eta squared. 95% CIs for ηp^2 computed using the MBESS package (Kelley, 2007). Effect sizes are described as large relative to Cohen's (1988) $\eta p^2 \geq .14$ convention. Observed power = 1.000 for all dimensions.

Social Engagement showed the largest effect size ($\eta p^2 = .837$, F (5, 145) = 148.80, $p < .001$). Behavioral Engagement was second ($\eta p^2 = .822$, F (5, 145) = 133.75, $p < .001$), followed by Cognitive Engagement ($\eta p^2 = .816$, F (5, 145) = 128.68, $p < .001$) and Emotional Engagement ($\eta p^2 = .775$, F (5, 145) = 99.97, $p < .001$). All four values exceed Cohen's (1988) conventional threshold for a large effect ($\eta p^2 = .14$) by a substantial margin, by a factor of approximately 5.5 to 6 in this sample; as noted above, this should be read as a description of the magnitude of within-cohort change over time rather than as a precise causal estimate.

In substantive terms, mean scores increased by approximately 1.2 to 1.6 scale points across all four dimensions. Social Engagement rose from a pre-intervention mean of M = 2.84 (Pre-intervention 1) to a post-intervention mean of M = 4.29 (Post-intervention 3). Behavioral Engagement increased from M = 2.78 to M = 4.41. Cognitive Engagement moved from M = 2.91 to M = 4.34, and Emotional Engagement from M = 2.75 to M = 4.16.

4.4. Pairwise Comparisons and the ITSD Temporal Pattern

Bonferroni-corrected pairwise comparisons are presented in Table 7. The pattern of results is consistent with the ITSD temporal signature described in Section 2.3, namely a stable pre-intervention baseline followed by a discontinuity at the intervention boundary and sustained post-intervention scores.

Across all four engagement dimensions, none of the twelve pre-intervention-to-pre-intervention comparisons reached statistical significance (0/12), consistent with a stable pre-intervention baseline. By contrast, all 36 pre-intervention-to-post-intervention comparisons reached statistical significance (36/36, $p < .001$, Bonferroni-corrected), indicating that every post-intervention occasion scored significantly higher than every pre-intervention occasion, for all four dimensions.

The comparison of Pre-intervention 3 with Post-intervention 1, capturing the period immediately surrounding commencement of the RWT intervention, was significant for all four dimensions (all $p < .001$), with mean differences ranging from 1.04 (Emotional Engagement) to 1.27 (Social Engagement). This discontinuity at the intervention boundary, occurring against a stable pre-intervention baseline, is consistent with (though does not, on its own, prove) an intervention-related change, in line with ITSD logic (Mark & Gamble, 2009).

Post-intervention-to-post-intervention comparisons yielded no significant results (0/12), suggesting that the gains observed at Post-intervention 1 were maintained through Post-intervention 2 and 3 without significant further increase or decrease within this cohort. This pattern is consistent with sustained engagement levels through the end of the term, though the study cannot speak to engagement beyond this window (see Section 7).

Table 7. Bonferroni-Corrected Pairwise Comparisons for All Engagement Dimensions: Key ITSD Contrasts

Comparison Type	Time Points	BE	CE	EE	SE	Total Significant
Pre-intervention to Pre-intervention	Pre1 vs Pre2	Ns	Ns	Ns	ns	0/4
	Pre2 vs Pre3	Ns	Ns	Ns	ns	0/4
	Pre1 vs Pre3	Ns	Ns	Ns	ns	0/4
Pre-intervention to Post-intervention	Pre1 vs Post1	✓	✓	✓	✓	4/4
	Pre1 vs Post2	✓	✓	✓	✓	4/4
	Pre1 vs Post3	✓	✓	✓	✓	4/4
	Pre2 vs Post1	✓	✓	✓	✓	4/4
	Pre2 vs Post2	✓	✓	✓	✓	4/4
	Pre2 vs Post3	✓	✓	✓	✓	4/4
	Pre3 vs Post1	✓	✓	✓	✓	4/4
	Pre3 vs Post2	✓	✓	✓	✓	4/4
	Pre3 vs Post3	✓	✓	✓	✓	4/4
Post-intervention to Post-intervention	Post1 vs Post2	Ns	Ns	Ns	ns	0/4
	Post1 vs Post3	Ns	Ns	Ns	ns	0/4
	Post2 vs Post3	Ns	Ns	Ns	ns	0/4

Note. Mean differences represent I minus J; negative values indicate J > I. $p < .001$, Bonferroni corrected. BE = Behavioral Engagement; CE = Cognitive Engagement; EE = Emotional Engagement; SE = Social Engagement. ns = not significant.

4.5. Polynomial Contrast Analysis: Decomposing the Engagement Trajectory

To characterize the temporal structure of the engagement trajectory in more detail, polynomial contrast analysis was conducted for the Social Engagement dimension, which showed the largest overall effect size. Results, presented in Table 8, indicate a temporal profile broadly consistent with a step-change interpretation.

The dominant component was Linear ($\eta p^2 = .958$, $F(1, 29) = 654.72$, $p < .001$), reflecting a consistent upward trend across the six occasions considered as a whole. The significant Cubic contrast ($\eta p^2 = .622$, $F(1, 29) = 47.69$, $p < .001$) is consistent with a pattern in which a relatively stable pre-intervention plateau is followed by a level shift around the intervention boundary and a subsequent re-stabilisation at a higher post-intervention level. The Quadratic contrast was non-significant ($\eta p^2 = .027$, $F(1, 29) = 0.80$, $p = .378$), which is inconsistent with a simple decelerating-growth (maturation-type) pattern, though it does not rule out maturation as a contributing factor more generally.

The significantly higher-order (fifth-order) contrast ($\eta p^2 = .711$, $F(1, 29) = 71.35$, $p < .001$) is consistent with a change in slope around the pre-/post-intervention boundary. The non-significant fourth-order contrast indicates an absence of marked oscillation within either phase, consistent with a relatively orderly pattern of change across the term, though, again, this descriptive pattern cannot by itself establish the underlying cause of the change.

Table 8. Tests of Within-Subjects Contrasts (Polynomial): Social Engagement Dimension

Source	Contrast	Type III Sum of Squares	df	Mean Square	F	Sig. (p)	Partial η^2	ITSD Interpretation
TIME	Linear	74.74	1	74.74	654.72	<.001	.958	Strong upward trend
	Quadratic	0.16	1	0.16	0.80	.378	.027	No intervention acceleration
	Cubic	5.69	1	5.69	47.69	<.001	.622	Intervention level shift
	Order 4	0.14	1	0.14	1.11	.302	.037	No oscillation
	Order 5	7.82	1	7.82	71.35	<.001	.711	Slope change after intervention

Note. SS = Type III Sum of Squares. Contrasts computed for Social Engagement (highest $\eta p^2 = .837$) as the reference dimension.

4.6. Qualitative Findings

Thematic analysis of the eight transcripts revealed four themes that broadly relate to the four quantitative dimensions of engagement, along with an additional theme pertaining to participants' worries about the intervention. Participant numbers (P1-P8) will be used throughout; P1-P3 had low levels of engagement pre-intervention, P4-P6 medium levels, and P7-P8 high levels (please see Section 3.2).

4.6.1. *Theme 1: 'I started doing more, and doing it on purpose' -- Behavioral engagement*

There were a few who felt like they had spent more time on writing assignments and deliberation in their actions. Participant 2 (P2) described the change from mere compliance to a more diligent process of spending more time on each assignment than they did in the beginning of the term, and not waiting for the last moment to write. Participant 5 (P5) felt that the weekly cycle was responsible for this kind of change, since the knowledge that the assignment would be reviewed by a fellow student and the journal entry to come made it harder for her to submit incomplete work.

4.6.2. *Theme 2: 'I started asking myself why' -- Cognitive engagement*

The secondary recurring theme included the development of planning and self-monitoring skills. For example, participant one (P1) reported shifting from writing drafts in a random fashion to outlining main points before actually starting writing, and checking afterwards whether all paragraphs adhere to the outlined plan. Participant four (P4) explicitly connected this change to reflective journals, claiming that while the question about the reason

for making some choices ("Why did you make this choice?") posed during the analysis phase seemed difficult initially, it was later integrated automatically into the writing process, not just done in the journal. Participant eight (P8) perceived the self-assessment rubric as helpful in discovering her typical mistakes, especially in terms of coherence between paragraphs.

4.6.3. *Theme 3: 'Feedback stopped feeling like an attack' -- Emotional engagement*

There were changes in how the participants experienced feedback and evaluation. Participant 3 (P3) remembered that at the beginning of the semester she felt like a "weak writer" when she got written feedback from the instructor, and sometimes she was reluctant to look through the feedback for one or two days. However, later on she felt eager to get feedback, as she had some questions regarding her writing. Participant 6 (P6) stated that there was a similar change in her attitude towards peer feedback – she felt uneasy having her classmates read her unfinished work at the beginning, but then she came to realize that peer feedback is "just another opinion to consider." Not all of the participants had such an absolute change in attitude: Participant 2 (P2) still felt some unease before talking to the instructor.

4.6.4. *Theme 4: 'Writing became something we did together' -- Social engagement*

The elements of the peer review process and the feedback dialogues figured prominently in the description of the social context of writing provided by the participants. Participant P5 pointed out the importance of the rotating peer-review process as the most prominent change from previous classes, when writing was just "between you and the teacher." Participant P7 said that the value of the dialogues with the instructor did not lie so much in the feedback itself but rather in the fact that it made it possible to immediately ask clarifying questions, which is impossible through written feedback only. Participant P1 noted that giving feedback as well as receiving it required thinking about writing from "the teacher's perspective," as described in Theme 2.

4.6.5. *Theme 5: Reservations and practical constraints*

There was another set of problems raised by the participants as well. The second participant (P2) and the eighth participant (P8) found the workload to be significant in the context of the amount of work expected in other modules since there were eight writing tasks and eight journal entries to complete in eight weeks. The third participant (P3) and sixth participant (P6) found that there was inconsistency in the quality of peer feedback, depending upon the pair, as some peers gave very generic feedback despite the availability of a structure provided through the rubric. The fourth participant (P4) mentioned that even though it had been assured to participants that the completion of interviews and questionnaires would not affect their grades, initially, it was difficult to separate the two roles of the lecturer.

4.7. Integrating Quantitative and Qualitative Findings

Table 9 presents a joint display integrating the quantitative effect sizes for each engagement dimension with the corresponding qualitative theme and illustrative participant accounts, together with an assessment of the degree of convergence between the two strands.

In general, there was convergence between the qualitative observations and the quantitative findings: effort, planning, comfort with feedback, and interactions with peers/instructors increased for all levels regardless of their baseline level, consistent with the direction of change seen in the quantitative findings (although not necessarily its magnitude). The qualitative data, on the other hand, provided evidence for two additional dimensions that were not clear from the quantitative findings: (a) individual variations in workload perceptions (Theme 5); and (b) partial relief from anxiety around feedback for at least one participant (P2, Theme 3).

Table 9. Joint Display of Quantitative and Qualitative Findings by Engagement Dimension

Engagement Dimension	Quantitative Pattern (ηp^2)	Qualitative Theme	Illustrative Account	Convergence
Behavioral	ηp^2 = .822; PRE1 M=2.78 → POST3 M=4.41	Theme 1: Increased, more deliberate effort	P2: increased time on task; P5: peer-review and journal deadlines reduced last-minute submission	Convergent
Cognitive	ηp^2 = .816; PRE1 M=2.91 → POST3 M=4.34	Theme 2: More planning and self-monitoring	P1: moved from unplanned writing to outlining; P4: analysis prompt became automatic by later weeks	Convergent
Emotional	ηp^2 = .775; PRE1 M=2.75 → POST3 M=4.16	Theme 3: Feedback experienced as less threatening (for most)	P3: feedback shifted from avoidance to anticipation; P2: residual nervousness persisted	Largely convergent, with one partial divergence (P2)
Social	ηp^2 = .837; PRE1 M=2.84 → POST3 M=4.29	Theme 4: Writing reframed as collaborative	P5: peer review as the most noticeable change; P7: dialogues enabled follow-up questions	Convergent
(cross-cutting)	Workload and context not captured by questionnaire items	Theme 5: Workload concerns and variable peer-feedback quality; partial discomfort with dual instructor/researcher role	P2 & P8: workload heavy; P3 & P6: variable peer-feedback quality; P4: difficulty separating instructor/researcher roles	Additional context (divergent from purely positive quantitative pattern)

5. Discussion

5.1. The Magnitude and Pattern of the Engagement Changes

As shown in Table 5, there are considerable changes in the self-reported levels of engagement over the span of eight weeks with partial eta-squared values of .775 – .837. It should be noted that effect sizes found in this study are significantly higher compared to typical effects sizes in second-language acquisition literature (.10 – .30 partial eta-squared values). Pre-intervention means (between 2.75 and 3.10 on a five-point scale) indicate that the group started the term with moderate levels of engagement which is consistent with the product-oriented and teacher-centered approach to teaching writing that is predominant in Ethiopia (Woldemariam, 2018). Post-intervention means (between 3.98 and 4.41) did not hit the ceiling point (5.00) which reduces the possibility that changes are caused by ceiling effects.

However, there are some caveats about the magnitude of the effects. Lack of the control group does not allow to exclude history, testing effects (frequent usage of the same questionnaire for six points) and social-desirability biases as possible causes of the effect. However, as mentioned above, measures described in Section 3.7 were taken to reduce the latter. Stable baseline and step changes from Pre-intervention 3 to Post-intervention 1 are consistent with an intervention-based explanation but do not confirm it by themselves. Stable baseline followed by step changes can be also associated with other events happening in the course of the term (changes in course material, time before mid-term exams, increased knowledge about the instructor and other students). Some qualitative data discussed later in the section provide additional but not conclusive evidence to intervention-based interpretation as participants related certain changes in their experiences to particular aspects of the RWT protocol.

5.2. Social Engagement and the Social Architecture of the Protocol

Social Engagement had the highest quantitative effect size ($\eta p^2 = .837$), among the four dimensions, whereas Theme 4 was evident in participants' descriptions as well, with some participants mentioning that this particular intervention created opportunities for collaborative writing compared to other courses they had attended before. This is consistent with the ideas proposed by Vygotsky (1978) in relation to the role of social interactions in cognitive development and also with the results reported by Bui and Nguyen (2026) and Teng and Huang (2026) concerning dialogue-infused reflection. The qualitative evidence suggests that such a result could be explained by the fact that both the peer review process and the feedback dialogues were important parts of the intervention. Yet, it might have been due to the difference between this and previous courses as suggested by the fact that one participant noted that writing had been "just between you and the teacher".

5.3. Emotional Engagement and Experiences of Feedback

However, the lowest (though still quite considerable) effect size was recorded for Emotional Engagement ($\eta p^2 = .775$). One possible way to interpret this result in light of qualitative data, specifically Theme 3 (Feedback), is that many participants described a two-phase process in which initially feedback was perceived as evaluative and even somewhat threatening for those participants who identified themselves as "weak writers" but gradually turned into a source of useful information (this process was evident for most participants but not all of them). These observations resonate to a certain degree with the notions of "error tolerance" and "feedback tolerance" discussed by Rad et al. (2026). Nonetheless, P2's continuous anxiety about participating in discussions during the whole term indicates that such a transition did not take place consistently and may partly account for a smaller quantitative effect size for Emotional Engagement.

One reason why Emotional Engagement may have developed at a slower pace than Behavioral and Social Engagement is possibly the idea that affective reactions to feedback depend on accumulated experience in multiple cycles, while behavioral engagement can change immediately in reaction to a certain structure of the task (a required journal entry).

5.4. Cognitive Engagement and Self-Regulation

However, the large effect size of Cognitive Engagement ($\eta p^2 = .816$) is consistent with the qualitative theme pertaining to planning and self-monitoring (Theme 2), since a number of participants reported moving from an initial state of unstructured writing towards planning, checking whether their draft matches the outline and using self-assessment rubrics to detect the repeated mistakes in their work. This is in line with the assumption that reflection can lead to self-regulation skills (Liang & Wu, 2025; Hegarty et al., 2026). Nevertheless, the current design is not sufficient to draw conclusions regarding whether those self-reports of changes in the strategies are associated with changes in actual behavior or performance.

5.5. Behavioral Engagement and Effort

The second-largest effect size ($\eta p^2 = .822$) and the highest post-intervention mean (Post-intervention 3, M = 4.41) were seen in Behavioral Engagement. The pre-intervention mean of Behavioral Engagement (Pre-intervention 1, M = 2.78) was one of the lowest of all four dimensions in terms of the first measure. The qualitative data (Theme 1) show a movement from minimum compliance-type of writing behavior into the effort put in on a consistent basis; several participants noted this change was due to certain structural aspects of the weekly schedule (submission of work to be reviewed by peers and journal entry). On the other hand, Theme 5 notes the increase in workload was seen as stressful for some participants, which is relevant to the future application of such a scheme (Section 7).

6. Conclusion

The present study explores the relationship between an eight-week structured intervention with Reflective Writing Tasks (RWTs) and changes in multidimensional engagement of EFL learners in one particular cohort at an Ethiopian university, and it includes accounts of participants' experiences throughout the period.

Substantial changes in time over all four dimensions of engagement (Behavioral, Cognitive, Emotional, and Social) occurred ($\eta p^2 = .775$ to .837, $p < .001$). In terms of temporal pattern, there was a stable baseline, a discontinuity at the intervention boundary, and continued post-intervention levels, which constitutes some evidence, albeit inconclusive, supportive of the intervention explanation. The biggest change was demonstrated by Social Engagement, and it appears that the peer-review and feedback-dialogue elements of the intervention protocol may have played an important role in this respect.

The findings are presented as a contextualized case study based on data collected from one particular cohort of students in one particular institution. We conclude that structured RWTs, as explained in Section 3.3, may be considered as an element of the teaching process in similar situations in Ethiopia, alongside with considerations of workload (Theme 5) and measures to address power dynamics, as explained in Section 3.7. While recognizing the limitations mentioned in Section 7 (the lack of control group and a small single-institution sample size), we do not offer any policy recommendations in terms of, for example, changing class size or mandatory introduction of structured RWTs in all EFL writing courses across the country. Instead, we recommend that the intervention protocol introduced above can be used in similar EFL writing courses at other institutions in Ethiopia, but in pilot programs, preferably involving a comparison group.

7. Limitations and Directions for Future Research

However, there are some limitations to this study that need to be taken into consideration when interpreting the results. First, there is no control or comparative group, which is inevitable for the design used, and so it makes it impossible to attribute any observed change in results to the RWT intervention as reliably as it could have been done with randomized and/or matched groups.

Second, this study has involved only one group of 30 third-year students from one university in Ethiopia and one course. So, it is only possible to generalize the findings to other student cohorts and institutions in the same way as if more than one group had been tested.

Third, the measure of quantifiable engagement used is a self-reporting scale that is subject to social desirability bias and response shift bias, whereby participants re-interpret what each scale point means due to their changed understanding of the concept being measured. Although certain measures (described in section 3.7) have been taken to reduce social desirability bias, this is still not enough to eliminate it.

Fourth, the dual position of the first author as an instructor of the course as well as the researcher poses a risk of introducing a power-dynamic bias with regard to instructor-mediated feedback dialogues and qualitative interviews. Though the interviews and the majority of dialogue monitoring sessions were held by the research assistant who was not involved in the process of grading, and the participants were informed repeatedly of a voluntary and non-evaluative nature of their participation in the research, there is a possibility that the participants adjusted their answers (P4 in Theme 5).

Fifth, the current study assessed the engagement during a period of eight weeks plus the short pre-intervention baseline all within one term. No follow-up measurement was made after the end of the term, let alone after the removal of the protocol, thus no information is available about the sustainability of the changes.

Sixth, the qualitative part of the research included only eight participants selected through a maximum variation sampling technique depending on the baseline engagement score. It must be noted that due to this method the obtained themes can be viewed as a set of possible experiences of this particular group, but not as representatives of the whole sample or all EFL learners in general.

In view of the above-mentioned limitations, further research should: (a) use randomized or matched comparison groups design (wherever it is possible), (b) repeat the procedure described in Section 3.3 in other Ethiopian universities, courses, and participant groups in order to check for the generalization of the findings, (c) explore mediating variables, including the quality and depth of feedback, depth of reflection (perhaps directly coded from journal entries instead of self-reporting), and writing self-efficacy; (d) make a follow-up measurement after the intervention period in order to check for the sustainability of the changes; (e) introduce a measure of the performance along with the measure of engagement in order to see whether the changes in the engagement are accompanied by changes in writing; and (f) consider the modification of the protocol in order to reduce the workload (Theme 5) and to make further checks against the power-dynamic problems of Theme 5 and Section 3.7, for example by conducting data collection and evaluation independently.

Data Availability Statement

The anonymized questionnaire data supporting the findings of this study (six measurement occasions, N = 30, identified by participant codes only) are openly available in the Mendeley Data repository (DOI: 10.17632/fpfxfp43zr.1; https://data.mendeley.com/datasets/fpfxfp43zr/1). The data may be reused with appropriate citation.

Acknowledgements

The authors thank the third-year English Language and Literature students at Jinka University who participated in this study, and the research assistant who administered the questionnaires and conducted the interviews. Institutional support from Debre Markos University and Jinka University is gratefully acknowledged.

Declaration of Conflict of Interest

The authors do not have any conflicts of interest.

Authors' contribution

Zewudu Paulos Mada led the study from conception to completion, including research design, data collection, analysis, and manuscript preparation. Sualih Mussa Akalu and Getinet Bilew Fente provided expert supervision, methodological and theoretical guidance, and critical feedback that strengthened the manuscript. All authors contributed to the final refinement of the paper, approved its publication, and share responsibility for its scholarly integrity.

References

Al-Hoorie, A. H. (2014). Engagement in language learning: A systematic review of 20 years of research methods and definitions. Language Teaching Research. https://doi.org/10.1177/13621688211001289

Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. SAGE Publications.

Bui, Q. T. T., & Nguyen, T. N. (2026). The role of Socratic dialogue in engaging EFL students in grammar lessons: Action research. LEARN Journal: Language Education and Acquisition Research Network, 19(1), 569-596.

Ceylan, A., & Korkmaz, H. (2026). The relationship between engagement levels of EFL learners and their academic achievement: A comparison across proficiency levels. Journal of Education, Theory and Practical Research, 12(1), 104-120. DOI: 10.38089/ekuad.2026.257

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.

Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process. D. C. Heath.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59-109. https://doi.org/10.3102/00346543074001059

Gibbs, G. (1988). Learning by doing: A guide to teaching and learning methods. Further Education Unit, Oxford Polytechnic.

Hansen, M., & Rossow, I. (2007). Adolescent gambling and problem gambling: Does the total consumption model apply? Journal of Gambling Studies, 24(2), 135-149. https://doi.org/10.1007/s10899-007-9083-5

Hegarty, A., Devenney, L., & Stephens, S. (2026). Behavioural, cognitive and affective factors of student engagement: A multi-modal model. International Journal of Teaching and Case Studies, 16(1), 85-110.

Hınız, G., & Çelik, Ö. (2025). A bibliometric and content analysis of student engagement research in English language teaching. Language Awareness, 34(1), 43-76. https://doi.org/10.1080/09658416.2023.2210946

Kelley, K. (2007). Confidence intervals for standardized effect sizes: Theory, application, and implementation. Journal of Statistical Software, 20(8), 1-24.

Kumari, J. M., & Prakasha, G. S. (2025). Reflective writing skills among pre-service teachers: A scoping review. International Journal of Evaluation and Research in Education, 14(5), 3801-3811.

Kurniawan, K. (2024). Enhancing 21st-century writing skills through reflective collaborative learning: A model based on critical thinking. Pakistan Journal of Life and Social Sciences, 22(2). https://doi.org/10.57239/PJLSS-2024-22.2.0023

Liang, H., & Wu, S. (2025). Dynamic pathways to academic engagement in university students: A mixed-methods study on growth mindset, grit, and academic self-efficacy with ecological momentary assessment. Frontiers in Psychology, 16, 1653578. https://doi.org/10.3389/fpsyg.2025.1653578

Mark, M. M., & Gamble, C. (2009). Experiments, quasi-experiments, and ethics. In D. M. Mertens & P. E. Ginsberg (Eds.), Handbook of social research ethics (pp. 198-213). SAGE Publications.

Marsden, E., & Torgerson, C. J. (2012). Single-group, pre- and post-test research designs: Some methodological concerns. Oxford Review of Education, 38(5), 583-616. https://doi.org/10.1080/03054985.2012.731208

Ministry of Education, Ethiopia. (2023). Education and training policy. Federal Democratic Republic of Ethiopia.

Pavlovich, K., Collins, E., & Jones, G. (2009). Developing students' skills in reflective practice: Design and assessment. Journal of Management Education, 33(1), 37-58. https://doi.org/10.1177/1052562907307640

Rad, H. S., Alipour, J., Jafarpour, A., & Hashemian, M. (2026). Integrating adaptive learning and expectancy-value theory of motivation to enhance L2 learners' writing achievement, error tolerance, and feedback tolerance. Research and Practice in Technology Enhanced Learning, 21, 040-040.

Rizqiani, D. A., & Astuti, M. T. (2025). Reflective writing as a pedagogical tool to foster critical awareness among non-English department students in Indonesian EFL contexts. Pedagogy: Journal of English Language Teaching, 13(2), 132-147.

Schon, D. A. (1983). The reflective practitioner: How professionals think in action. Basic Books.

Schon, D. A. (1987). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. Jossey-Bass.

Teng, M. F., & Huang, J. (2026). Assessing ChatGPT feedback for EFL learners' engagement and writing performance. International Journal of Applied Linguistics. https://doi.org/10.1111/ijal.12606

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Woldemariam, C. (2018). Academic writing challenges in Ethiopian higher education: A case study. Ethiopian Journal of Education and Sciences, 14(1), 1-20.