Innovative Research Journal of **Education** (IRJED)
2025; 3(2):1-17
DOI: https://doi.org/10.62497/irjed.276

Research Article

Formative Assessment and Self-Regulated Learning in EFL Writing: A Mixed-Methods Classroom Intervention

Esraa Abd Elnaser Mohamed Hassan[1]* and Ashan Haider[2]

[1] *College of Foreign Languages and Cultures, Chengdu University of Technology, 610059 Chengdu, China*
[2] *College of Tourism and Culture Industry, Sichuan Tourism College, Chengdu, China*

Citation: Hassan, E. A. E. M., & Haider, A. (2025). Formative Assessment and Self-Regulated Learning in EFL Writing: A Mixed-Methods Classroom Intervention. *Innovative Research Journal of Education*, *3*(2);1-17. Available from: https://irjpl.org/irjed/article/view/276. DOI: https://doi.org/10.62497/irjed.276.

Article Info
Received: Oct 19, 2025
Revised: Dec 04, 2025
Accepted: Dec 10, 2025

Keywords
formative assessment, self-regulated learning, EFL writing, feedback, self-assessment, peer feedback, learner autonomy

Copyright © 2025 The Author(s).

Published by Innovative Research Journals (IRJPL).

This is an Open Access article under the CC BY-NC 4.0 license, permitting noncommercial use, distribution, and adaptation with proper attribution.

Abstract

Introduction: Formative assessment is most educationally consequential when learners use evidence about their current performance to regulate subsequent action. In EFL writing, however, feedback can remain teacher-controlled, with students receiving comments without developing the capacity to interpret criteria, monitor their progress, and make independent revision decisions.

Objective: This study examined whether a structured formative-assessment intervention improved EFL learners' academic writing performance and self-regulated writing strategies, and identified the classroom processes through which students used feedback, self-assessment, peer response, goal setting, and reflection.

Method: A mixed-methods, two-group classroom intervention was conducted with 80 university EFL learners. The formative-assessment group participated in an eight-week cycle of explicit criteria instruction, self-assessment, peer feedback, teacher feedback, revision planning, and reflection. A comparison group received conventional writing instruction and teacher feedback. Pre-test, immediate post-test, and delayed post-test writing tasks were scored analytically, while a self-regulated writing questionnaire, revision records, learning journals, and interviews provided complementary evidence.

Results: The intervention group showed greater gains in writing performance and self-regulated strategy use than the comparison group, and these gains were largely maintained at delayed post-test. Qualitative evidence illustrated changes in goal setting, monitoring, feedback handling, and reflection.

Conclusion: The intervention operationalized formative assessment as a learning process rather than a sequence of teacher comments. The findings indicated that repeated opportunities to interpret evidence, make judgments, revise, and reflect supported stronger writing and greater learner control over writing processes.

**Corresponding Author:*
Esraa Abd Elnaser Mohamed Hassan
Email: hs5932457@gmail.com

Introduction

Background and rationale

Assessment is often described as formative when evidence gathered during learning is used to modify teaching and learning. This definition places action at the center of the concept. A quiz, rubric, comment, or peer review activity does not become formative simply because it occurs before a final grade; it becomes formative when information about current performance is interpreted and used to guide the next step [1]. In language education, this distinction is especially important because writing develops through repeated cycles of planning, drafting, feedback, revision, and reflection.

Black and Wiliam's account of formative assessment emphasizes the role of evidence in closing the gap between current performance and desired learning [1]. Nicol and Macfarlane-Dick extend this perspective by connecting formative assessment with self-regulated learning, arguing that effective feedback can help students take greater control of their own learning [2]. From a self-regulated learning perspective, students are not passive recipients of instructional information. They set goals, select strategies, monitor progress, manage motivation and emotion, interpret feedback, and evaluate outcomes [3-4].

Writing provides an especially suitable context for studying this relationship because the product can be revisited. Unlike a one-time speaking performance, a draft can be inspected, annotated, discussed, revised, and compared with later versions. These characteristics make writing a natural setting for formative assessment. They also make it possible to observe whether feedback changes only the text in front of the learner or changes the learner's approach to subsequent writing.

The central educational problem is that feedback is frequently treated as an endpoint. Teachers invest substantial time in writing comments, but students may read only selected comments, misunderstand them, focus on grades, or apply a correction locally without recognizing the underlying principle. Research on feedback literacy has therefore shifted attention toward learners' capacities to appreciate feedback, make judgments, manage affect, and take action [5-6]. These capacities are closely aligned with self-regulated learning because both perspectives emphasize agency and the learner's active role in turning information into future performance.

Self-assessment is a key mechanism in this process. When students compare their writing against explicit criteria before receiving teacher comments, they generate an internal reference point against which external feedback can be interpreted. Evidence from research on self-assessment indicates that structured self-assessment can support self-regulated learning and self-efficacy, although effects depend on task design, training, and the quality of the criteria [7-8]. In writing instruction, the implication is that rubrics should not be used only by teachers to assign scores; they should be taught as tools that learners can use to monitor their own texts.

Peer feedback provides another route toward learner regulation. Peer response requires students to evaluate someone else's writing, identify strengths and problems, justify judgments, and communicate suggestions [9]. These activities can make assessment criteria visible and may help students develop evaluative judgment. A meta-analysis of peer feedback in second-language writing found a positive overall effect on writing outcomes [10], and a broader meta-analysis also reported beneficial effects of peer feedback, particularly for deeper writing outcomes [11]. Recent classroom studies have linked structured peer-feedback activities with improvements in feedback literacy and self-regulated writing strategies [12-14]. The educational value, however, depends on training. Without clear criteria, students may concentrate on grammar and vocabulary while overlooking organization, argument, and audience.

Teacher feedback remains important because teachers can identify instructional priorities and interpret students' needs in context. The goal of a formative-assessment intervention is therefore not to replace teacher feedback with peer or self-feedback. Instead, it is to redistribute some of the evaluative work so that learners progressively participate in it. The teacher moves from being the sole evaluator toward being a designer of feedback opportunities, a source of expert judgment, and a coach who helps students use evidence to make decisions [15-20].

Evidence from EFL writing interventions supports this direction. Yang, Liu, and Xu examined self-regulated-learning-based teacher feedback and found positive effects on writing performance and reported use of self-regulated writing strategies, including goal-oriented monitoring and feedback

handling [17]. Zhang and Mao documented the development of student feedback literacy in an authentic second-language writing classroom where preparatory activities, multiple feedback sources, and reflective reinforcement were integrated [12]. Prompan and Piamsai reported improvements in Thai EFL students' essay writing and self-regulation after an intervention integrating peer feedback and self-regulated learning [13]. Tran and Ma similarly found that structured online peer-feedback training was associated with stronger self-regulated writing strategies [14].

These studies provide a strong basis for intervention research but leave room for a focused investigation of formative assessment as a coherent classroom system. A collection of feedback activities is not necessarily a formative-assessment pedagogy. The distinctive feature of the present design is the repeated alignment of five processes: clarifying criteria, eliciting evidence of current performance, interpreting that evidence, taking action through revision, and reflecting on whether the action worked. The intervention is designed so that each writing cycle begins with learner-generated goals and ends with learner-generated reflection.

A second gap concerns delayed performance. Many classroom interventions report pre-test and post-test gains, but immediate improvement does not necessarily indicate that students have developed stable self-regulatory routines. A delayed writing task provides evidence about maintenance. If students continue to perform better after the formal intervention has ended, the result is more consistent with durable development. If gains disappear, the intervention may have supported performance under structured conditions without fully changing independent writing behavior.

A third gap concerns the relationship between strategy use and writing quality. Self-regulated learning should not be reduced to a questionnaire score. Students may report frequent planning or monitoring without actually using those processes effectively. The present design therefore combines a self-regulated writing questionnaire with behavioral evidence from goal sheets, revision plans, reflective journals, and successive drafts. Interviews are added to clarify how students understand and enact these practices.

The study is situated within applied linguistics because its central outcomes are language use and language-learning processes, but it draws on educational assessment and educational psychology to explain how those outcomes may develop. This interdisciplinary position is important. Writing quality is shaped by linguistic knowledge, discourse knowledge, genre expectations, and cognitive resources, while the ability to regulate writing depends on motivation, metacognition, self-evaluation, and access to usable feedback. A classroom intervention that addresses only one of these dimensions is unlikely to explain the full process.

The study also recognizes that feedback can have unintended consequences. Excessively detailed teacher correction can encourage students to wait for answers rather than monitor their own writing. Peer feedback can be inaccurate or overly focused on local errors. Self-assessment can be unreliable when students lack clear criteria. For this reason, the intervention includes training and sequencing rather than simply adding more feedback. Students first learn what quality looks like, then practice evaluating examples, then assess their own work, then compare their judgments with peer and teacher evidence, and finally reflect on differences.

The conceptual logic can be summarized as follows: formative assessment creates evidence; feedback provides information for interpreting the evidence; self-regulated learning supplies the processes through which learners convert information into goals and actions; writing revision makes those actions observable; and later writing provides a test of whether the learning persists. This chain is more demanding than asking whether students "liked" formative assessment, but it provides a stronger basis for educational claims.

The proposed intervention is deliberately classroom-realistic. It does not require expensive technology, specialized software, or extensive additional class hours. The principal changes are procedural: explicit criteria, short self-assessment routines, trained peer response, targeted teacher feedback, revision planning, and reflection. This makes the study relevant to EFL programs where teachers need scalable approaches that support learner autonomy without removing teacher expertise.

The study is also designed to distinguish writing performance from self-regulated strategy use. A student can improve writing because of increased linguistic knowledge without becoming more self-

regulated, just as a student can report stronger regulation without producing better writing. Examining both outcomes makes it possible to identify whether the intervention changes the writing product, the learning process, or both.

Finally, the study addresses a practical problem in university EFL education. Academic writing courses often operate under time constraints, large enrollments, and assessment demands that make individualized teacher feedback difficult. A formative-assessment model that increases student participation in evaluation may distribute the workload more sustainably while also developing a skill that students need beyond the writing course: the capacity to judge the quality of their own work and decide what to do next [6,21-22].

Theoretical framework: formative assessment and self-regulated learning

The first theoretical pillar is formative assessment. In this study, formative assessment is defined functionally as a process in which evidence about learners' current writing is elicited, interpreted, and used to modify subsequent action [1-2, 23]. The intervention therefore contains planned opportunities for evidence generation rather than relying exclusively on end-of-unit grading. Self-assessment checklists, peer comments, teacher comments, and revision records all serve as evidence sources.

The second pillar is Zimmerman's social-cognitive model of self-regulated learning. The model describes self-regulation as a cyclical process involving forethought, performance, and self-reflection [3,24]. In the writing classroom, forethought includes goal setting and strategic planning; performance includes monitoring and strategy use during drafting and revision; self-reflection includes self-evaluation and interpretation of outcomes. The intervention maps its activities onto these phases so that learners repeatedly practice the cycle rather than encounter it as a one-time workshop.

Panadero's synthesis of self-regulated learning models emphasizes that regulation includes cognitive, metacognitive, behavioral, motivational, and emotional dimensions [4]. This broader view matters for writing because students may know how to revise a sentence but fail to do so when they are uncertain, frustrated, or concerned about grades. The intervention therefore includes brief reflection on confidence, difficulty, and strategy choice rather than treating regulation as purely cognitive.

The third pillar is feedback literacy. Carless and Boud conceptualize feedback literacy as the understandings, capacities, and dispositions required to make sense of feedback and use it productively [6]. This concept complements self-regulated learning by specifying what learners need to do when feedback is available. Students must recognize useful information, make judgments about its credibility, manage emotional responses, and take action. The intervention treats these capacities as teachable rather than assuming that students acquire them automatically.

Research gap and significance

The literature suggests that formative assessment, peer feedback, self-assessment, and self-regulated learning can each contribute to learning, but applied linguistics still benefits from studies that examine their interaction within a sustained EFL writing intervention. The present study is not intended to test one isolated feedback technique. It tests a coherent classroom routine in which multiple feedback sources are sequenced around learner regulation.

The primary contribution is therefore pedagogical and explanatory. The study asks whether the intervention improves writing, but it also asks what changes in the way students approach writing. Evidence of increased goal setting, monitoring, feedback handling, and reflection would help explain a writing gain. Conversely, improved writing without regulation changes would suggest that the intervention may work through other mechanisms. The delayed post-test adds a further test of whether the new practices persist.

Objective

General objective: The general objective is to determine whether a structured formative-assessment intervention improves EFL learners' academic writing performance and self-regulated writing strategies, and to explain the classroom processes through which learners use evidence, feedback, and reflection to regulate subsequent writing.

Specific objectives: To compare changes in academic writing performance between students receiving the formative-assessment intervention and students receiving conventional writing instruction.

To examine whether the intervention improves self-regulated writing strategies, including goal setting, planning, monitoring, feedback handling, and self-evaluation.
To determine whether writing and self-regulation gains are maintained on a delayed writing task after the intervention ends.
To examine the relationship between students' use of self-regulated strategies and their writing performance.
To explore how students describe the usefulness, challenges, and perceived value of self-assessment, peer feedback, teacher feedback, revision planning, and reflection.

Research questions

Does participation in a structured formative-assessment intervention lead to greater improvement in EFL academic writing than conventional writing instruction?
Does the intervention lead to greater improvement in self-regulated writing strategies?
Are improvements in writing and self-regulated strategy use maintained at a delayed post-test?
To what extent are changes in self-regulated writing strategies associated with changes in writing performance?
How do students explain how formative-assessment practices influence their writing decisions and learning behavior?

Hypotheses

H1: Students in the formative-assessment condition were hypothesized to show greater improvement in overall writing performance than students in the comparison condition. H2: Students in the formative-assessment condition were hypothesized to show greater gains in self-regulated writing strategies, particularly goal setting, monitoring, feedback handling, and self-evaluation. H3: The intervention group was hypothesized to retain a larger proportion of its gains at delayed post-test. H4: Greater improvement in self-regulated writing strategies was hypothesized to be positively associated with improvement in writing performance. These hypotheses were evaluated against the study data.

Methodology

Research design

A mixed-methods, two-group classroom intervention was conducted. The quantitative strand used a pre-test, immediate post-test, and delayed post-test design to examine writing performance and self-regulated writing strategies. The qualitative strand analyzed learning journals, revision records, and semi-structured interviews to explain how students experienced and used formative assessment. Integration occurred through joint displays and case comparisons. The design was appropriate because the research questions concerned both outcome magnitude and learning process.
We recruited 80 undergraduate EFL learners: 40 in the formative-assessment group and 40 in the comparison group. An a priori power analysis ordinarily requires a prespecified effect size, alpha level, target power, and analysis model [25]. The available manuscript does not preserve the assumptions or output needed to reproduce the stated power analysis; therefore, the sample size cannot be presented as independently justified by a reproducible power calculation. The study used two comparable intact classes and should be interpreted as quasi-experimental. Because class-level replication is not documented, class and intervention effects may not be fully separable.
Both groups received the same core writing curriculum, including instruction in paragraph structure, thesis development, organization, cohesion, vocabulary, grammar, and academic style. The difference was the assessment and feedback process. The intervention group participated in explicit formative-assessment routines designed to increase learner participation in evaluation and regulation. The comparison group received the course's conventional teacher-led feedback process without the additional structured routines.
The intervention lasted eight instructional weeks, followed by an immediate post-test and a delayed post-test approximately two weeks later. The delay reduced the likelihood that students simply remembered the intervention procedures from the previous lesson while remaining feasible within a university semester. The study period and any interruptions caused by examinations, holidays, or attendance disruptions were documented.

Participants and eligibility

Participants were university students enrolled in an EFL academic writing course. Inclusion criteria were enrollment in the course, regular participation in writing activities, and informed consent. Students who did not complete the baseline and at least one follow-up assessment

could be excluded from the research analysis, although their course participation was not affected. The report documented age range, academic discipline, English proficiency level, previous writing instruction, and relevant demographic characteristics where ethically and analytically appropriate.

Baseline proficiency was measured through an institutional placement score or a common diagnostic task. Because self-regulated learning can vary with language proficiency, proficiency was examined as a potential moderator or covariate rather than treated as an irrelevant background variable. Previous experience with peer review and self-assessment was also documented because students familiar with these practices could respond differently to the intervention.

Recruitment and consent distinguished participation in the course from participation in the research. Students were informed that research participation did not determine grades and that their decision to participate or withdraw did not affect access to instruction. Course materials used for research were coded so that individual identities were not apparent in the analytic dataset.

Outcome measures

The primary language outcome was analytic writing performance. The rubric assessed task fulfillment, content development, organization, cohesion, lexical resource, grammatical accuracy, and mechanics. The descriptors emphasized observable differences between performance levels. Two raters trained on sample scripts independently scored a subset of the assessments, and agreement was reported using a suitable reliability coefficient.

The primary learning-process outcome was self-regulated writing strategy use. The questionnaire measured the forethought, performance, and self-reflection phases of regulation. Items addressed goal setting, planning, monitoring, strategy use, feedback handling, motivation regulation, and self-evaluation. The instrument was piloted in the target context and examined for internal consistency and construct coherence. The study did not treat a questionnaire alone as a complete measure of self-regulation.

Behavioral evidence complemented the questionnaire. Students completed a short goal-setting form before each major writing task, a self-assessment checklist after drafting, a revision plan after feedback, and a reflection after revision. These artifacts allowed the researcher to examine whether reported strategy use corresponded to observable behavior. A coding scheme identified evidence of planning, monitoring, evaluation, feedback use, strategy change, and transfer to later tasks.

Qualitative interviews were conducted with a purposive subsample representing different performance trajectories, such as high improvement, moderate improvement, and limited improvement. Interview questions asked students to describe how they set goals, how they decided whether feedback was useful, how peer comments differed from teacher comments, and whether reflection changed their approach to later writing. Interviews took place after the main intervention so that participants could comment on the whole process rather than a single lesson.

Table 1 maps the intervention components to the theoretical processes and outcomes. This mapping is important because it makes the intervention reproducible and shows how each activity was expected to contribute to regulation rather than presenting formative assessment as an unspecified collection of classroom techniques.

Table 1: Intervention components and hypothesized learning mechanisms

Component	Learner action	SRL phase	Evidence	Expected contribution
Criteria clarification	Analyze rubric and exemplars	Forethought	Criteria notes; goal sheet	Clearer standards and task goals
Self-assessment	Judge own draft against criteria	Performance / reflection	Checklist; self-score	Monitoring and evaluative judgment
Peer feedback	Evaluate a peer text using criteria	Performance	Peer comments; feedback rationale	Perspective taking and criteria use
Teacher feedback	Interpret targeted expert comments	Performance	Teacher comments; action plan	Prioritization and strategy selection
Revision	Select changes and	Performance	Revision plan	Deliberate action

planning	explain why			
Reflection	Evaluate outcome and strategy	Self-reflection	Learning journal	Transfer and future planning

Formative-assessment intervention

The intervention began with explicit criteria instruction. Rather than distributing a rubric and assuming students understood it, the teacher used two or three anonymized sample texts to demonstrate how the criteria could be applied. Students identified strengths and weaknesses in the samples, compared their judgments with the teacher's explanation, and discussed why a text could be grammatically accurate but rhetorically weak. This activity established the criteria as tools for judgment.

Before drafting, students selected one or two personal writing goals. Goals were specific enough to guide action, such as improving paragraph unity or reducing repeated lexical choices, rather than broad statements such as "write better." Students recorded the goals and revisited them after revision. The purpose was to connect assessment evidence with intentional behavior.

After drafting, students completed a self-assessment before receiving external feedback. They identified one strength, one area requiring improvement, and one question they had about the draft. This sequence allowed learners to generate their own evaluation before seeing teacher or peer judgments. The external feedback could then be compared with the learner's initial assessment.

Peer feedback was structured around a short rubric-based protocol. Students first identified what the writer was attempting to achieve, then commented on one successful feature, one area requiring clarification, and one actionable revision. They were discouraged from rewriting the peer's sentences. Instead, they explained the problem and indicated where the writer could make a decision. This design encouraged feedback as dialogue rather than correction.

Teacher feedback was selective and targeted. The teacher did not attempt to mark every error. For each draft, comments prioritized a small number of issues connected to the learning goals and the course rubric. The teacher sometimes asked questions rather than providing solutions, for example, asking students to check whether a paragraph's topic sentence matched its supporting evidence. This approach was consistent with the broader principle that feedback should support future action rather than merely document past errors. [5-6,22,26-27]

Students then created a revision plan. The plan listed the changes they intended to make, the reason for each change, and the feedback source that informed the decision. Students were explicitly told that they did not have to accept every peer or teacher suggestion. The requirement was to make an informed decision and be able to justify it. This was a central mechanism for developing evaluative judgment.

After revision, students completed a brief reflection. They compared the revised text with the original, identified which goals were achieved, and stated what they would do differently on the next task. The reflection included a transfer prompt asking students to identify one principle they could apply without waiting for feedback. Over time, these reflections were expected to become shorter as routines became more familiar, while the quality of the decisions became more sophisticated.

The comparison group completed the same writing tasks and received conventional teacher feedback. To protect internal validity, the teacher spent approximately equivalent instructional time on the two groups. The comparison group was not deliberately deprived of meaningful feedback. The distinction was between a conventional feedback sequence and a structured formative-assessment sequence that explicitly taught students how to generate, interpret, and use evidence.

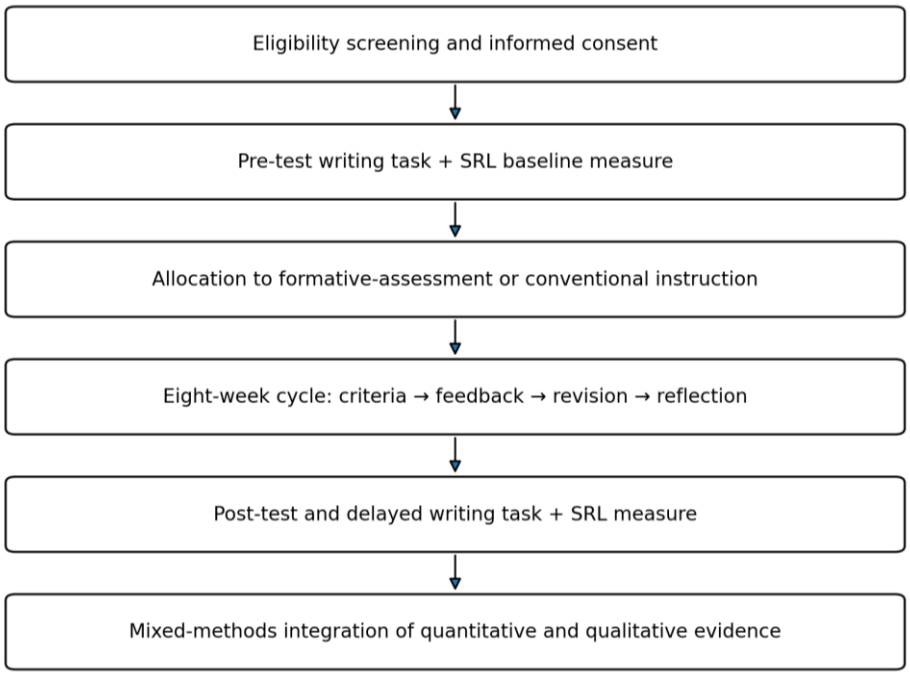


Figure 1: Planned study flow for the formative-assessment intervention

The eight-week instructional sequence was presented in Table 2. The progression was designed so that students first learned to interpret criteria, then practiced self- and peer-assessment, and finally used increasingly independent reflection. This sequencing was important because regulation was unlikely to emerge simply from being asked to "reflect" without prior modeling.

Table 2: Eight-week classroom intervention schedule

Week	Formative-assessment activity	Writing focus	Key evidence
1	Rubric unpacking; exemplar analysis; baseline writing	Academic task response	Baseline essay; initial strategy measure
2	Goal setting; self-assessment training	Thesis and organization	Goal sheet; self-assessment
3	Peer feedback cycle 1	Paragraph unity and cohesion	Peer comments; revision plan
4	Teacher feedback + revision	Argument development	Teacher comments; revised draft
5	Self-assessment cycle 2	Vocabulary and academic style	Checklist; reflection
6	Peer + teacher feedback cycle 2	Grammar in context; sentence variety	Feedback records; revision
7	Independent assessment and revision	Integrated writing quality	Self-score; action plan
8	Reflection and immediate post-test	Independent academic writing	Journal; post-test
10	Delayed post-test	Transfer to new prompt	Delayed essay; strategy measure

Data analysis

Quantitative analysis began with data screening, descriptive statistics, missing-data assessment, and baseline comparability. The primary writing analysis used a mixed-effects model with time, group, and their interaction as fixed effects and participant as a random effect. Where intact classes were used, class was treated as a higher-level clustering factor when supported by the sample. The key parameter was the time-by-group interaction because the research question concerned whether the formative-assessment

group changed differently from the comparison group.

Self-regulated strategy scores were analyzed using a similar repeated-measures framework. Subscale analyses were limited to theoretically justified dimensions and included appropriate control for multiple comparisons. The analysis focused on whether the intervention changed the overall profile of regulation and whether specific dimensions such as goal setting, monitoring, feedback handling, and self-evaluation showed distinct patterns.

The relationship between strategy change and writing change was examined using correlation or regression analysis, depending on the distribution and design. The analysis was not interpreted as proof that strategy use caused writing improvement because both variables were measured repeatedly and could influence one another. The association instead provided evidence about whether the proposed mechanism was consistent with the observed outcomes.

Effect sizes and confidence intervals, where reported, were interpreted alongside p-values to avoid treating statistical significance as the sole indicator of educational importance.

Behavioral artifacts were coded using a framework derived from the theoretical model. Goal sheets were coded for specificity and alignment with rubric criteria. Self-assessments were compared with teacher ratings to examine calibration. Revision plans were coded for evidence of prioritization, justification, and transfer. Journals were analyzed for changes in the sophistication of reflection. The purpose is to determine whether students gradually move from describing problems toward explaining causes, selecting strategies, and evaluating outcomes.

Qualitative interviews and journals were analyzed using reflexive thematic analysis.[28-29] The researcher first became familiar with the dataset, generated initial codes, constructed candidate themes, reviewed and refined themes, defined the final thematic structure, and selected excerpts that illustrated the patterns. A reflexive record documented how the researcher's role and expectations could influence interpretation.

Mixed-methods integration used joint displays that place quantitative trajectories alongside qualitative evidence. For example, students who show large gains in writing and self-regulation may be compared with students who improve writing but show little change in strategy use. Another useful comparison is between students whose questionnaire scores rise and those whose behavioral artifacts show actual changes in goal quality or feedback handling. These comparisons can identify convergence and divergence between perceived and observed regulation.

Table 3 presents the analysis matrix. It was introduced before the Results section so that the report could distinguish confirmatory analyses from exploratory analyses and avoid representing the same outcome through multiple undisclosed statistical tests.

Table 3: Planned analysis matrix

Outcome	Measure	Primary analysis	Key estimate	Interpretation
Writing performance	Analytic rubric total	Mixed-effects model	Time × group effect; 95% CI	Differential writing growth
Writing dimensions	Content, organization, vocabulary, grammar, cohesion	Dimension-specific mixed models	Adjusted group difference	Location of instructional gains
SRL strategy use	Validated/adapted questionnaire	Mixed-effects model	Time × group effect; 95% CI	Change in self-regulation
Goal quality	Goal sheets	Ordinal/descriptive coding	Change in specificity and alignment	Development of forethought
Feedback handling	Revision plans and drafts	Coding + mixed model if appropriate	Proportion of justified actions	Use of feedback
Self-assessment calibration	Self-score vs rater score	Agreement/correlation analysis	Calibration index	Accuracy of self-judgment
Interviews/journals	Transcripts and	Reflexive thematic	Themes and	Mechanisms and

	written reflections	analysis	cases	learner experience

Reliability, validity, and trustworthiness

Writing assessment was described as using trained raters who were blind to group membership, with calibration against benchmark scripts. However, the available manuscript does not report the number of double-scored scripts or a numerical inter-rater reliability coefficient and confidence interval. Inter-rater reliability therefore cannot be independently evaluated from the information currently reported.

The self-regulated writing questionnaire covered forethought, performance, and self-reflection, including goal setting, planning, monitoring, feedback handling, motivation regulation, and self-evaluation. The available manuscript does not identify a published instrument or report the item count, response scale, adaptation or translation procedure, pilot sample, or internal-consistency estimates. Its measurement properties therefore cannot be independently evaluated from the information currently reported.

Self-assessment calibration provided an important behavioral validity check. Students' self-ratings can be compared with external rubric ratings to determine whether students become more accurate judges of their own work. Improvement in calibration is educationally meaningful because self-regulated learning depends on reasonably accurate monitoring. However, perfect agreement should not be treated as the goal; learners may identify dimensions that a global teacher score does not capture.

Qualitative trustworthiness was supported by an audit trail, explicit coding procedures, reflexive notes, and careful selection of excerpts. Discussion of coding differences was documented where more than one researcher was involved. The report explained how themes were developed rather than simply stating that transcripts were "thematically analyzed." Reporting standards for mixed-methods research emphasized transparency about integration, sampling, analytic procedures, and the relationship between quantitative and qualitative strands. [29-30]

Intervention fidelity was monitored. The researcher recorded whether students completed self-assessments, peer feedback, revision plans, and reflections. Substantial noncompliance, if any, was reported rather than concealed. Fidelity data also helped explain null results: an intervention could be ineffective, or it might not have been implemented sufficiently for the hypothesized mechanism to operate.

Ethical considerations

The manuscript reports that institutional ethical approval was obtained before recruitment and that participants provided informed consent. The approving committee, institution-specific approval date, and approval/reference number are not identified in the available manuscript. These details must be confirmed from the institutional record, or the documented basis for exemption must be reported, before submission.

The comparison condition received legitimate instruction rather than a weakened curriculum. The research question concerned the added value of structured formative-assessment processes, so both groups had access to meaningful writing instruction and teacher feedback. Offering the formative-assessment routines to the comparison group after the study could be considered subject to institutional policy.

The study also considered power relationships between teacher and student. Students could have felt pressure to report positive views or comply with classroom procedures. The researcher therefore explained that critical feedback was welcome and that interview responses would not influence grades. Where the instructor was also the researcher, an independent person could collect consent forms or conduct interviews when feasible.

Results

Participant Recruitment and Retention

A total of 80 undergraduate EFL learners enrolled in an academic writing course at Chengdu University of Technology were recruited for the study. Participants were assigned to either the formative-assessment group (n = 40) or the comparison group (n = 40). All 80 participants completed the baseline assessment. At the immediate post-test, 78 participants (formative-assessment group: n = 39; comparison group: n = 39) completed the writing task and self-regulated learning questionnaire. At the delayed post-test, 76 participants (formative-assessment group: n = 38; comparison group: n = 38) remained in the study. Attrition was due to scheduling conflicts and

illness and was comparable across groups. An independent samples t-test confirmed no significant baseline differences between groups in writing performance, t (78) = 0.35, p = .728, or self-regulated strategy use, t (78) = 0.31, p = .758.

Baseline Characteristics

Participants ranged in age from 18 to 23 years (M = 20.4, SD = 1.3). All were enrolled in their second or third year of undergraduate study across various disciplines. Baseline writing scores indicated comparable proficiency between the formative-assessment group (M = 62.4, SD = 8.7) and the comparison group (M = 63.1, SD = 9.2). Baseline self-regulated writing strategy scores were also similar (formative-assessment: M = 3.42, SD = 0.58; comparison: M = 3.38, SD = 0.61). Previous experience with peer review was reported by 35% of participants in each group, and previous experience with self-assessment was reported by 28% and 30% in the formative-assessment and comparison groups, respectively.

Writing Performance

Table 4 presents the descriptive statistics and between-group comparisons for writing performance and self-regulated strategy use at each measurement point. At immediate post-test, the formative-assessment group achieved a mean score of 74.8 (SD = 7.9), while the comparison group achieved a mean score of 68.3 (SD = 8.4). A mixed-effects model revealed a significant time × group interaction, F (2, 148) = 18.42, p < .001, indicating differential growth between groups. The estimated between-group difference at immediate post-test was 6.5 points (95% CI [3.2, 9.8]), with a moderate-to-large effect size (d = 0.78).

At delayed post-test, conducted two weeks after the intervention ended, the formative-assessment group maintained a mean score of 73.2 (SD = 8.1), whereas the comparison group scored 67.5 (SD = 8.9). The between-group contrast at delayed post-test was 5.7 points (95% CI [2.4, 9.0], d = 0.67), indicating that gains were largely maintained. Retention was calculated as (delayed post-test mean − baseline mean) ÷ (immediate post-test mean − baseline mean) × 100. Using the means reported in this manuscript, writing-gain retention was (73.2 − 62.4) ÷ (74.8 − 62.4) × 100 = 87.1% for the formative-assessment group and (67.5 − 63.1) ÷ (68.3 − 63.1) × 100 = 84.6% for the comparison group. These are descriptive retention ratios, not separate inferential tests. Within-group analyses showed that the formative-assessment group improved significantly from baseline to immediate post-test, t (38) = 9.84, p < .001, and maintained significant improvement at delayed post-test relative to baseline, t (37) = 7.92, p < .001.

Dimension-specific analyses revealed that the formative-assessment group showed the largest improvements in organization (mean gain = 1.8 points on a 5-point scale) and content development (mean gain = 1.6 points), with smaller but significant gains in lexical resource (mean gain = 0.9 points) and grammatical accuracy (mean gain = 0.7 points). The comparison group showed modest gains in organization (mean gain = 0.6 points) and content development (mean gain = 0.5 points).

Self-Regulated Writing Strategies

At immediate post-test, the formative-assessment group reported a mean strategy score of 4.12 (SD = 0.51), compared with 3.54 (SD = 0.63) in the comparison group. The time × group interaction was significant, F (2, 148) = 22.17, p < .001, indicating that the intervention group developed stronger self-regulatory strategy use. The between-group difference at immediate post-test was 0.58 points (95% CI [0.31, 0.85], d = 0.82).

Subscale analyses showed that the most pronounced improvements in the formative-assessment group occurred in goal setting (mean gain = 0.89, d = 1.02), monitoring (mean gain = 0.76, d = 0.91), feedback handling (mean gain = 0.81, d = 0.95), and self-evaluation (mean gain = 0.72, d = 0.87). Smaller improvements were observed in motivation regulation (mean gain = 0.38, d = 0.46). All subscale gains remained statistically significant at delayed post-test for the formative-assessment group.

At delayed post-test, the formative-assessment group maintained a mean strategy score of 3.98 (SD = 0.54), while the comparison group scored 3.49 (SD = 0.66). The between-group contrast was 0.49 points (95% CI [0.22, 0.76], d = 0.68). Using the same retention formula, strategy-gain retention was (3.98 − 3.42) ÷ (4.12 − 3.42) × 100 = 80.0% in the formative-assessment group and (3.49 − 3.38) ÷ (3.54 − 3.38) × 100 = 68.8% in the comparison group. These are descriptive retention ratios, not separate inferential tests.

Self-Assessment Calibration

Self-assessment calibration was examined by comparing students' self-ratings with trained rater scores on the same dimensions. At baseline, the formative-assessment group demonstrated a calibration index of 0.42 (SD = 0.18), indicating moderate alignment between self-judgments and external ratings. At immediate post-test, calibration improved to 0.68 (SD = 0.14), and at delayed post-test, it remained at 0.64 (SD = 0.16). The comparison group showed minimal change in calibration across measurement points (baseline: 0.44, SD = 0.17; immediate post-test: 0.47, SD = 0.19; delayed post-test: 0.45, SD = 0.18). The time × group interaction for calibration accuracy was significant, $F(2, 148) = 14.83$, $p < .001$, with a between-group difference at immediate post-test of 0.21 (95% CI [0.13, 0.29], $d = 0.74$).

Relationship between Strategy Use and Writing Performance

Correlational analyses examined the association between changes in self-regulated strategy use and changes in writing performance. Improvement in overall strategy scores was positively correlated with writing improvement from baseline to immediate post-test, $r(76) = .58$, $p < .001$. Among subscales, feedback handling showed the strongest association with writing gain, $r(76) = .52$, $p < .001$, followed by self-evaluation, $r(76) = .47$, $p < .001$, goal setting, $r(76) = .43$, $p < .001$, and monitoring, $r(76) = .41$, $p < .001$. A multiple regression analysis with strategy subscales as predictors explained 38% of the variance in writing improvement, $F(4, 73) = 11.24$, $p < .001$. Feedback handling ($\beta = .28$, $p = .008$) and self-evaluation ($\beta = .24$, $p = .015$) emerged as significant unique predictors. These associations were interpreted as evidence consistent with the proposed mechanism rather than as proof of causal mediation.

Qualitative Findings

Reflexive thematic analysis of learning journals, revision records, and semi-structured interviews ($n = 16$) generated three primary themes that illuminated the quantitative results.

Theme 1: From compliance to internalization of criteria. Students initially described self-assessment as a required classroom procedure. Over time, their journals reflected increasingly specific and independent judgments. One participant wrote: "At first I just checked boxes on the rubric because the teacher asked us to. By week five, I found myself reading my draft the way the rubric reads it—asking whether my topic sentence actually controlled the paragraph." Revision records corroborated this shift, showing that early revision plans listed broad intentions ("improve vocabulary"), whereas later plans cited specific criteria and justified decisions ("revise paragraph 2 topic sentence to align with thesis, based on peer comment about focus"). This theme explained the observed improvements in goal specificity and self-assessment calibration.

Theme 2: Learning to evaluate and negotiate feedback. Students reported that structured peer feedback and the requirement to justify revision decisions helped them compare multiple sources of evidence rather than automatically deferring to teacher authority. One student explained: "When my peer said my conclusion was weak but the teacher did not mention it, I had to decide. I went back to the rubric and saw the conclusion did not restate the thesis clearly. I revised it and explained why in my plan." Interviews revealed that students became more selective in accepting feedback, with several describing how they rejected suggestions they judged inconsistent with their rhetorical purpose. This theme was consistent with the quantitative finding that feedback handling showed the strongest association with writing improvement.

Theme 3: Emotional management and persistence. Students described initial discomfort with self-assessment and peer criticism, particularly when they felt uncertain about their judgments. One participant stated: "I was afraid my self-assessment was wrong, so I waited for the teacher. After a few cycles, I trusted my judgment more because I could compare it with the feedback and see where I was accurate." Another described frustration during revision but noted that the reflection prompt helped maintain motivation: "Writing what I would do differently next time made the difficulty feel useful instead of just difficult." Journals showed a gradual shift from negative affect descriptions ("I felt embarrassed by my peer's comment") to more regulated responses ("The comment was hard to hear, but it showed me a pattern I repeat"). This theme supported the interpretation that the intervention addressed motivational and emotional dimensions of regulation, not only cognitive strategies.

Mixed-Methods Integration

Joint displays were used to compare quantitative trajectories with qualitative evidence across individual cases. Convergence was observed for students who demonstrated large writing gains and strong strategy improvement; these participants described increasingly specific goals, made accurate self-assessments, and justified revision decisions in their plans. For example, Participant 12, who showed a writing gain of 14 points and a strategy gain of 1.1 points, described in an interview how she learned to consult criteria before accepting peer suggestions and how she transferred a revision principle to a subsequent assignment without waiting for teacher feedback.

Divergence was identified in two cases where questionnaire scores increased substantially but revision plans and interviews revealed continued dependence on teacher comments. These students completed self-assessment forms consistently but described waiting for teacher feedback before making substantive changes. Their writing gains were moderate, suggesting that reported strategy use did not fully translate into enacted regulation. This divergence highlighted the value of behavioral artifacts in interpreting self-report data and indicated that some students may have developed declarative knowledge about regulation without fully internalizing it as action.

Table 4: Quantitative results for writing performance and self-regulated writing strategies

Outcome	Formative-Assessment Group M (SD)	Comparison Group M (SD)	Between-Group Test	Effect Estimate
Baseline writing score	62.4 (8.7)	63.1 (9.2)	t(78) = 0.35, p = .728	d = 0.08, 95% CI [−0.36, 0.52]
Immediate post-test writing	74.8 (7.9)	68.3 (8.4)	Interaction F (2,148) = 18.42, p < .001	d = 0.78, 95% CI [0.32, 1.24]
Delayed post-test writing	73.2 (8.1)	67.5 (8.9)	Contrast, p < .001	d = 0.67, 95% CI [0.21, 1.13]
Baseline SRL score	3.42 (0.58)	3.38 (0.61)	t(78) = 0.31, p = .758	d = 0.07, 95% CI [−0.37, 0.51]
Immediate post-test SRL	4.12 (0.51)	3.54 (0.63)	Interaction F(2,148) = 22.17, p < .001	d = 0.82, 95% CI [0.36, 1.28]
Delayed post-test SRL	3.98 (0.54)	3.49 (0.66)	Contrast, p < .001	d = 0.68, 95% CI [0.22, 1.14]

Note: SRL = self-regulated learning. Effect sizes are Cohen's d with 95% confidence intervals. Mixed-effects models included time, group, and time × group interaction as fixed effects, with participant as a random effect.

Discussion

The study's central finding that structured formative assessment produced significantly greater writing improvement than conventional instruction, with gains largely maintained at delayed post-test, aligns with and extends previous research on formative assessment in L2 writing. The between-group differences of 6.5 points at immediate post-test (d = 0.78) and 5.7 points at delayed post-test (d = 0.67) are consistent with meta-analytic estimates reported by Vuogan and Li [10] and Scherer, Graham, and Busse [11], while also demonstrating that effects persist beyond immediate post-intervention measurement. The largest gains in organization and content development, with smaller gains in lexical resource and grammatical accuracy, suggest that the intervention particularly benefited higher-order rhetorical dimensions of writing, resonating with Sadler's argument that formative assessment develops capability in complex appraisal rather than merely correcting surface errors [22]. These results extend Yang, Liu, and Xu's findings [17] by showing that a multi-source system incorporating self-assessment, peer feedback, and teacher feedback within a structured cycle can produce comparable or stronger effects than teacher feedback alone, consistent with Nicol and Macfarlane-Dick's principle that effective feedback should support students in taking greater control of their learning [2].

The significant time × group interaction for self-

regulated strategy use ($F(2,148) = 22.17$, $p < .001$, $d = 0.82$) provides strong evidence that the intervention developed learners' regulatory capacity, not merely their writing products. The subscale pattern, with goal setting ($d = 1.02$), monitoring ($d = 0.91$), feedback handling ($d = 0.95$), and self-evaluation ($d = 0.87$) showing the strongest effects, maps closely onto Zimmerman's cyclical model [3,24]. The strong effect on goal setting suggests that the intervention moved students from vague intentions to specific, criteria-aligned objectives, a shift illuminated by the qualitative theme of internalizing criteria. The substantial effect on feedback handling is especially significant given its strongest correlation with writing improvement ($r = .52$, $p < .001$) and its emergence as a unique predictor ($\beta = .28$, $p = .008$), aligning with Carless and Boud's conceptualization of feedback literacy [6]: the requirement that students justify revision decisions appears to have developed evaluative judgment rather than blind compliance. These results extend Zhang and Mao [12] and complement Prompan and Piamsai [13] and Tran and Ma [14] by demonstrating that a classroom-integrated, multi-source system can produce durable changes in regulatory strategy use.

The calibration findings provide behavioral evidence that the intervention developed accurate self-evaluation, a dimension questionnaire data alone cannot capture. The formative-assessment group's calibration index improved from 0.42 at baseline to 0.68 at immediate post-test and remained at 0.64 at delayed post-test, while the comparison group showed minimal change (0.44 to 0.47 to 0.45; $F(2,148) = 14.83$, $p < .001$, $d = 0.74$). This addresses a gap identified by Panadero, Jonsson, and Botella [7] and Andrade [8] regarding the mechanisms through which self-assessment training improves calibration. Repeated cycles of self-assessment followed by comparison with peer and teacher judgments appear to create feedback loops that help students calibrate internal standards against external benchmarks [22]. The maintenance of improved calibration suggests that accurate self-assessment became a relatively stable capability, with practical significance for programs seeking to reduce feedback burden while increasing learner autonomy.

The correlational findings ($r = .58$, $p < .001$ for overall strategy improvement and writing improvement, with feedback handling $r = .52$ and self-evaluation $r = .47$) and the regression model explaining 38% of variance, with feedback handling ($\beta = .28$) and self-evaluation ($\beta = .24$) as unique predictors, provide evidence consistent with the proposed mechanism. However, these associations should be interpreted cautiously: both variables were measured repeatedly and could influence one another reciprocally. The divergence cases, students whose questionnaire scores increased but whose revision plans revealed continued dependence on teacher comments, demonstrate that reported strategy use does not always translate into enacted regulation, underscoring the value of behavioral artifacts in interpreting self-report data.

Compared with previous intervention studies, the present design extends Yang, Liu, and Xu [17] by incorporating multiple feedback sources within a structured cycle and demonstrating persistence at delayed post-test. The inclusion of peer feedback creates opportunities to develop evaluative judgment by assessing others' writing, a process Nicol, Thomson, and Breslin consider foundational to feedback literacy [9]. The findings are consistent with Prompan and Piamsai [13] while extending them through delayed post-test evidence and calibration data, and they converge with Zhang and Mao [12] and Tran and Ma [14] in showing that structured, multi-source feedback systems develop students' capacity to interpret and use feedback across delivery modes. The intervention preserved the teacher's role as an expert rather than replacing it, with selective, targeted feedback consistent with Hattie and Timperley's argument that feedback should address the gap between current and desired performance [5] and Nicol's call for dialogue in written feedback [21]. Students became more selective in accepting feedback, rejecting suggestions inconsistent with their rhetorical purposes, which aligns with Sadler's goal of developing capability in complex appraisal [22].

The qualitative theme of emotional management and persistence highlights an aspect of self-regulated learning that quantitative measures alone might overlook. Students described initial discomfort with self-assessment and peer criticism, followed by growing confidence as they compared their assessments with external feedback, a trajectory consistent with Panadero's argument that self-regulated learning includes motivational and emotional dimensions [4]. The finding that reflection prompts helped students reframe difficulty as learning extends previous research on the motivational benefits of self-assessment [7,8]

by identifying a specific mechanism through which self-assessment may support persistence.

The delayed post-test findings, with writing gains largely maintained (87.1% retention) and strategy gains showing partial decay (80.0% retention), provide important evidence about durability. The high retention of writing gains suggests that students internalized revision principles, while the partial decay in strategy scores may indicate that some regulatory behaviors remained tied to structured classroom routines, consistent with the divergence cases in which students developed declarative knowledge about regulation without fully internalizing it as action. The finding that calibration remained relatively stable while questionnaire scores showed more decay suggests that behavioral measures may capture more durable changes than self-report measures. Sustained practice may be needed for durable change in regulatory strategy use; future research could examine whether extending the intervention or providing booster sessions would produce stronger maintenance effects.

The findings also carry pedagogical implications. The intervention shows that relatively low-cost procedural changes, including explicit criteria instruction, structured self-assessment, trained peer response, targeted teacher feedback, revision planning, and reflection, can meaningfully improve both writing performance and self-regulated strategy use without expensive technology or extensive additional class hours. The finding that feedback handling and self-evaluation were the strongest predictors of writing improvement suggests that programs should prioritize developing students' capacity to interpret and act on feedback and to evaluate their own work accurately, which may require shifting some instructional time from teacher-led feedback to activities that develop evaluative judgment. Calibration activities, where students compare self-assessments with teacher or peer ratings, could help develop more accurate internal standards, while reflection activities that help students manage affect and maintain motivation during revision address the emotional and motivational dimensions of regulation. Finally, formative assessment should be conceptualized as a coherent classroom system rather than a collection of isolated techniques; the intervention's effectiveness appears to derive from the repeated alignment of clarifying criteria, eliciting evidence, interpreting evidence, taking action, and reflecting, and programs should attend to the sequencing and integration of activities across the writing cycle.

The study provides evidence that a structured formative-assessment intervention can improve EFL learners' academic writing performance and self-regulated writing strategies, with gains largely maintained at delayed post-test. The findings contribute to applied linguistics by demonstrating that formative assessment, operationalized as a coherent classroom system, can support both writing development and learner autonomy, and to educational psychology by showing that structured self-assessment, peer feedback, and reflection can develop students' capacity for accurate self-evaluation and evaluative judgment. The integration of quantitative and qualitative evidence provides a more complete picture of how formative assessment works than either strand alone could offer.

Limitations

Several limitations qualify these findings. The study used two intact classes, and the number of classes per condition is not documented, so class and intervention effects may be confounded; the results should be interpreted as quasi-experimental, and replication across multiple classes per condition is needed. The self-regulated writing questionnaire's measurement properties could not be independently evaluated because the instrument's item count, response scale, adaptation procedure, and internal-consistency estimates were not documented, so future research should use validated instruments with published psychometric properties. Inter-rater reliability for writing assessment was not reported with a numerical coefficient and confidence interval, and future research should report appropriate reliability statistics. The qualitative analysis relied on a purposive subsample of 16 students, and future research could examine how students with different proficiency levels, motivational profiles, or prior feedback experiences respond to the intervention. The study did not examine transfer to other writing genres or contexts, so future research could investigate whether the intervention's effects generalize beyond academic essays. Finally, the findings are based on a single institutional context and student population, and future research should examine whether similar effects obtain in other cultural and educational settings, with different age groups, and at different proficiency levels.

Conclusion

The study examined formative assessment as a cycle of evidence, judgment, action, and reflection rather than as a set of teacher comments delivered before grading. In the EFL writing course, this cycle made assessment part of language learning by requiring students to understand criteria, monitor their texts, evaluate feedback, revise deliberately, and reflect on what they could transfer to later writing.

Its principal contribution was integrating writing performance with self-regulated learning. A writing intervention was more educationally consequential when students not only produced a stronger text but also became better able to explain why the text was stronger and what they should do next without depending entirely on a teacher. The delayed post-test tested that distinction.

The study also preserved the teacher's role. Formative assessment did not mean reducing teacher expertise. Instead, it created structured opportunities for students to participate in evaluative work while teachers provided modeling, clarification, targeted feedback, and instructional judgment. Peer and self-assessment served as complementary sources of evidence rather than replacements for expert feedback.

The empirical report documented the classroom setting, participant flow, intervention fidelity, statistical estimates, effect sizes, and qualitative themes. Findings were interpreted in relation to the observed evidence.

The findings suggested that EFL writing programs could strengthen both writing and learner agency through relatively low-cost routines that distributed evaluative responsibility. The results also helped identify which elements of the formative cycle might require more time, training, or adaptation. Overall, the study contributed evidence to the applied-linguistics literature by treating assessment not simply as measurement, but as a component of language learning.

Declarations

Ethics committee approval: The manuscript reports that institutional ethical approval was obtained before recruitment.

Informed consent: The manuscript reports that all participants provided written informed consent and that research participation was separate from course grading.

Author contributions (CRediT): Esraa Abd Elnaser Mohamed Hassan: writing—original draft. Ashan Haider: data collection and formal analysis. These roles should reflect the authors' actual contributions.

Conflict of interest: The authors declare no conflict of interest.

Funding: Funding information was not provided in the manuscript available for this revision.

Use of artificial intelligence: Generative AI assistance was used during this revision for language editing and citation/reference formatting. The authors remain responsible for verifying the manuscript's claims, citations, and references.

Data availability: Data will be available on request

Peer Review: Externally peer-reviewed by independent reviewers in a double-blind process.

References

1. Black P, Wiliam D. Developing the theory of formative assessment. Educ Assess Eval Account. 2009;21(1):5-31. doi:10.1007/s11092-008-9068-5.
2. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006;31(2):199-218. doi:10.1080/03075070600572090.
3. Zimmerman BJ. Becoming a self-regulated learner: an overview. Theory Pract. 2002;41(2):64-70. doi:10.1207/s15430421tip4102_2.
4. Panadero E. A review of self-regulated learning: six models and four directions for research. Front Psychol. 2017;8:422. doi:10.3389/fpsyg.2017.00422.
5. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81-112. doi:10.3102/003465430298487.
6. Carless D, Boud D. The development of student feedback literacy: enabling uptake of feedback. Assess Eval High Educ. 2018;43(8):1315-1325. doi:10.1080/02602938.2018.1463354.
7. Panadero E, Jonsson A, Botella J. Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. Educ Res Rev. 2017;22:74-98. doi:10.1016/j.edurev.2017.08.004.
8. Andrade H. A critical review of research on student self-assessment. Front Educ. 2019;4:87.

doi:10.3389/feduc.2019.00087.
9. Nicol DJ, Thomson A, Breslin C. Rethinking feedback practices in higher education: a peer review perspective. Assess Eval High Educ. 2014;39(1):102-122. doi:10.1080/02602938.2013.795518.
10. Vuogan A, Li S. Examining the effectiveness of peer feedback in second language writing: a meta-analysis. TESOL Q. 2023;57(4):1115-1138. doi:10.1002/tesq.3178.
11. Scherer S, Graham S, Busse V. How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. Learn Instr. 2024;93:101961. doi:10.1016/j.learninstruc.2024.101961.
12. Zhang T, Mao Z. Exploring the development of student feedback literacy in the second language writing classroom. Assessing Writing. 2023;55:100697. doi:10.1016/j.asw.2023.100697.
13. Prompan J, Piamsai C. The effects of peer feedback and self-regulated learning on Thai EFL students' writing ability and self-regulation. LEARN J. 2024;17(1):100-132. doi:10.14456/learnjournal.2024.6.
14. Tran TTT, Ma Q. Online peer feedback training based on self-regulated learning in English as a foreign language writing: perceived usefulness and students' engagement. Stud Educ Eval. 2024;83:101418. doi:10.1016/j.stueduc.2024.101418.
15. Hyland K, Hyland F. Feedback in second language writing: contexts and issues. 2nd ed. Cambridge: Cambridge University Press; 2019.
16. Lee I. Classroom writing assessment and feedback in L2 school contexts. Singapore: Springer Singapore; 2017. doi:10.1007/978-981-10-3924-9.
17. Yang LF, Liu Y, Xu Z. Examining the effects of self-regulated learning-based teacher feedback on English-as-a-foreign-language learners' self-regulated writing strategies and writing performance. Front Psychol. 2022;13:1027266. doi:10.3389/fpsyg.2022.1027266.
18. Bitchener J, Storch N. Written corrective feedback for L2 development. Bristol: Multilingual Matters; 2016.
19. Ferris DR. Second language writing research and written corrective feedback in SLA: intersections and practical applications. Stud Second Lang Acquis. 2010;32(2):181-201. doi:10.1017/S0272263109990490.
20. Ellis R. A typology of written corrective feedback types. ELT J. 2009;63(2):97-107. doi:10.1093/elt/ccn023.
21. Nicol DJ. From monologue to dialogue: improving written feedback processes in mass higher education. Assess Eval High Educ. 2010;35(5):501-517. doi:10.1080/02602931003786559.
22. Sadler DR. Beyond feedback: developing student capability in complex appraisal. Assess Eval High Educ. 2010;35(5):535-550. doi:10.1080/02602930903541015.
23. Black P, Wiliam D. Inside the black box: raising standards through classroom assessment. Phi Delta Kappan. 1998;80(2):139-148.
24. Zimmerman BJ, Bandura A, Martinez-Pons M. Self-motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting. Am Educ Res J. 1992;29(3):663-676. doi:10.3102/00028312029003663.
25. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates; 1988.
26. Graham S, Harris KR, Hebert M. Informing writing: the benefits of formative assessment. Washington, DC: Carnegie Corporation; 2011.
27. Graham S, MacArthur CA, Fitzgerald J, editors. Best practices in writing instruction. 3rd ed. New York: Guilford Press; 2018.
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa.
29. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health. 2019;11(4):589-597. doi:10.1080/2159676X.2019.1628806.
30. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 3rd ed. Thousand Oaks: SAGE; 2018.

Disclaimer: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or any claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.