Ilmiy tadqiqotlar va ularning yechimlari jurnali
Aoka guvohnoma № 323759 ISSN: 3060-4818

DEVELOPING EFL LEARNERS' AI FEEDBACK LITERACY: A PEDAGOGICAL FRAMEWORK FOR CRITICAL EVALUATION AND SELF-REGULATED WRITING

Norboyeva Muxlisa Alimardon qizi
MA in Teaching English to Speakers
of Other Languages (TESOL)
https://doi.org/10.5281/zenodo.23333064

Abstract

Generative artificial intelligence can provide English as a foreign language (EFL) learners with immediate writing feedback, yet access to suggestions does not ensure informed revision or durable learning. This conceptual article develops a pedagogical framework for teaching learners to interpret, evaluate, and use AI feedback while retaining responsibility for their writing. A targeted narrative synthesis integrates feedback literacy, evaluative judgement, AI literacy, and self-regulated learning with research on automated and generative feedback. The synthesis identifies five interdependent capabilities: task-aligned feedback elicitation; interpretation and prioritisation; evidence-based evaluation and calibrated trust; management of agency, affect, and ethical responsibility; and revision, reflection, and transfer. These capabilities are organised into a recursive instructional cycle linking goal setting, feedback requests, verification, justified revision, and independent application. A six-week teaching sequence, an illustrative revision record, and performance-based assessment procedures operationalise the framework. The central proposition is that educational value depends on the quality of learners' decisions rather than the proportion of suggestions accepted. The framework is theoretically grounded but has not been empirically validated. Its contribution is to specify teachable practices and testable outcomes through which AI-assisted revision may support increasingly independent EFL writing.

Keywords: AI feedback literacy; EFL writing; generative artificial intelligence; evaluative judgement; self-regulated learning; critical evaluation; learner agency; writing pedagogy

1 Introduction

Writing in a foreign language requires learners to coordinate ideas, linguistic resources, genre conventions, and audience expectations. Feedback can help them identify discrepancies between intended and achieved meaning, but its usefulness depends on what learners do with it. Generative AI expands the opportunities for receiving feedback: a learner can request an explanation, ask for a more focused comment, or explore an alternative wording while drafting. These opportunities also expand the decisions learners must make. A suggestion can be grammatically plausible yet distort the argument, introduce unsupported information, or move the text away from its intended audience.

The resulting pedagogical problem concerns the relationship between assistance and learning. When an improved sentence is inserted into a draft, the written product may become more effective. Whether the learner understands that improvement, can justify it, and can produce a comparable sentence independently remains a separate question. Instruction therefore needs to make the reasoning

between receiving a suggestion and revising a text visible. Without that intermediate work, writing assessment risks attributing tool-supported performance to a capability the learner has not yet developed.

Feedback literacy provides a useful starting point. Carless and Boud (2018) identify appreciating feedback, making judgements, managing affect, and taking action as interconnected capacities. Their account places learners within the feedback process as interpreters and decision makers. In an AI-mediated environment, this role includes assessing the source itself. Learners must consider whether a comment is accurate, relevant to the task, and compatible with the meaning they intend to communicate. A confident explanation cannot substitute for such evaluation.

Research already recognises AI feedback literacy as a distinct area of inquiry. Liu and Deris (2025) developed and examined the construct with university EFL learners, linking it to feedback uptake and motivational appraisals. Accordingly, the present article does not claim to introduce the term. It addresses the instructional question of how these capacities can be taught, observed in actual revision decisions, and connected to later writing without AI assistance. That emphasis is particularly relevant when confidence in using a tool exceeds the learner's ability to assess its recommendations.

The purpose is to develop a pedagogical framework for critical evaluation and self-regulated EFL writing. Its objectives are to integrate relevant theories, identify teachable capabilities, design a classroom sequence, and specify assessment procedures that distinguish revision quality from independent learning. Three questions guide the analysis:

RQ1. What capabilities enable EFL learners to evaluate AI-generated writing feedback critically?

RQ2. How can these capabilities be embedded in a recursive cycle of self-regulated writing?

RQ3. What classroom evidence can demonstrate informed feedback decisions and transfer to independent writing?

The proposed framework is intended primarily for tertiary EFL writing instruction, with adaptations for learners at approximately B1–B2 proficiency. This is a proposed application context, not a description of participants in a new study. The framework remains open to other proficiency levels and genres, provided that verification tasks and linguistic demands are adjusted. Its organising principle is that accepting, adapting, rejecting, or deferring a suggestion should follow a reasoned judgement about the writing problem and the learning goal.

2 Theoretical foundations

2.1 Feedback literacy and evaluative judgement

Feedback literacy concerns the learner's capacity to work productively with information about performance. Molloy et al. (2020) extend this discussion through a learning-centred framework encompassing feedback purposes, information seeking, judgement, emotions, and action for future work. Their analysis supports treating literacy as a set of enacted practices. A learner who recognises the value of feedback but cannot use a comment to plan a revision has not yet demonstrated the capability that instruction seeks to develop.

Evaluative judgement explains how learners determine whether work meets relevant standards. Tai et al. (2018) connect understanding quality with making decisions about one's own and others' work. In EFL writing, quality includes linguistic accuracy, coherence, evidence, genre, and communicative purpose. These dimensions can conflict. Replacing a simple expression with a more elaborate one may

increase formality while reducing precision or readability. A rubric and contrasting exemplars help learners discuss such trade-offs, although neither removes the need for situated judgement.

The AI context requires this judgement to encompass both the suggested product and the process used to obtain it. Bearman et al. (2024) argue that generative AI makes the development of evaluative judgement especially significant. The present framework translates that concern into a classroom requirement: learners explain what a proposed change improves, what it might damage, and what evidence supports their decision. The proposed comparison activities initially conceal source labels so that learners assess feedback quality before considering its origin.

2.2 AI literacy and the status of generated feedback

General AI literacy includes competencies for interacting with and critically evaluating AI systems (Long & Magerko, 2020). Ng et al. (2021) organise the field around understanding, application, evaluation and creation, and ethical issues. For writing instruction, these broad competencies need to be situated within an actual task. Knowing that AI can make errors is insufficient unless a learner can identify which claim requires checking and choose an appropriate verification method.

Large language models generate responses that can be fluent without being reliable. Kasneci et al. (2023) discuss educational opportunities alongside limitations involving accuracy, bias, and responsible use. Consequently, a generated explanation is treated here as a hypothesis about the writing problem. It may direct attention to a useful issue, but it does not independently establish a grammar rule, validate a factual statement, or authenticate a citation. Asking the same system to confirm its answer can support clarification; it does not provide independent corroboration.

AI feedback literacy is therefore related to, but more task-specific than, general AI literacy. It also differs from prompt competence. A well-constructed request can improve the relevance of the response, while leaving its correctness uncertain. The distinction matters pedagogically because learners may become proficient at eliciting sophisticated comments before they develop the linguistic or disciplinary resources necessary to assess them.

2.3 Self-regulated learning and learner engagement

Zimmerman's (2002) account of self-regulated learning links forethought, performance, and self-reflection. Panadero (2017) situates this cycle within a broader family of models that address cognitive, motivational, emotional, and behavioural regulation. Applied to AI-assisted writing, forethought involves identifying a writing goal and deciding what assistance is appropriate; performance involves monitoring feedback and revision; reflection involves evaluating the result and adjusting future strategies.

Nicol and Macfarlane-Dick (2006) connect formative feedback with learners' existing ability to assess their work and generate internal feedback. This supports placing self-assessment before the AI request. A learner first identifies a suspected problem, then uses external feedback to test that diagnosis. Such sequencing provides a basis for comparing personal judgement with the tool's comments and reduces the likelihood that an unsolicited list of corrections determines the entire revision agenda.

Engagement research adds an important qualification. Zhang and Hyland's (2018) study of two learners over a semester illustrates the role of engagement in using teacher and automated feedback. Ranalli (2021) further identifies trust as consequential in learners' interaction with automated feedback.

These studies concern earlier automated writing systems; their findings inform pedagogical design but cannot establish the performance of contemporary generative models. Together, the theoretical strands suggest that useful assistance requires both evaluative resources and opportunities to practise regulation within meaningful writing tasks.

3 Methodology

3.1 Conceptual design and source selection

This article adopts a conceptual design combining theory synthesis with pedagogical model development. Jaakkola (2020) distinguishes conceptual approaches that integrate theoretical perspectives and clarify relationships between constructs. Here, the analytic objective is to connect feedback evaluation with the regulation of writing, rather than estimate an intervention effect. No learners were recruited, no classroom experiment was conducted, and no original quantitative results are reported.

A targeted narrative search was completed on 13 October 2026. Search phrases combined "AI feedback literacy," "EFL writing," "ChatGPT feedback," "automated writing feedback," "evaluative judgement," and "self-regulated learning." Publisher websites, university repositories, and author-deposited manuscripts were consulted. The reference set contains 20 scholarly publications, including foundational theory, research reviews, learner studies, a comparison of feedback quality, and a recent simulation study. Foundational publications were retained for their theoretical relevance, without imposing a recent-publication cutoff.

Sources were selected for direct relevance to the research questions, identifiable publication details, and access to an abstract or full text sufficient to substantiate the particular claim used. Authors, titles, publication years, and DOI identifiers were checked against publisher or institutional records. Abstract-only access supported only claims stated in the abstract. Secondary summaries were used for discovery, not as substitutes for the original research. The procedure was purposive and does not constitute a systematic review or claim exhaustive coverage of the field.

3.2 Analytic procedure and limits

The synthesis used four analytic steps. First, it identified constructs explaining feedback interpretation, judgement, agency, and regulation. Second, it examined writing studies for opportunities, difficulties, and limits of inference. Third, it translated those insights into observable learner actions and aligned them with writing phases. Fourth, it checked the proposed activities against assessment evidence: every intended capability needed a corresponding task or record through which it could be observed.

Published findings, interpretive connections, and proposed classroom practices are distinguished throughout. Earlier automated systems, generative AI, and simulated writers are not treated as interchangeable evidence. Selection bias and uneven access to full texts limit the breadth of the synthesis. The outcome is a theoretically informed framework requiring empirical validation, not a validated scale or an established causal model.

4 Results of the conceptual synthesis

4.1 Pedagogical needs identified in existing research

The literature supports a conditional account of AI-assisted writing. Shi and Aryadoust's (2024) systematic review examines 83 automated written feedback studies published between 1993 and 2022. It provides a substantial background for understanding instructional implementation, but its coverage largely precedes widespread generative AI use. Its relevance lies in showing that feedback technology has a pedagogical history; conclusions about newer tools require evidence specific to their interactive and generative characteristics.

Steiss et al. (2024) compared 200 sets of human feedback with 200 sets of ChatGPT feedback on the same secondary-school essays. Trained human evaluators performed better on the assessed dimensions other than alignment with criteria. This was a comparison of feedback quality, not a test of EFL learners' long-term development. Nevertheless, it identifies a reason to teach learners to examine accuracy, actionable guidance, and prioritisation rather than equate an apparently rubric-based response with a sufficient evaluation.

Mahapatra (2024) reported positive academic writing outcomes in a mixed-method intervention with undergraduate ESL students in India. The study's six-hour intervention and limited genre coverage restrict conclusions about durability and broader applicability. Its findings support investigating AI feedback within structured instruction; they do not establish that unsupervised access produces independent writing development in every EFL setting.

More directly, Yan and Zhang (2024) investigated four EFL writers' behavioural, cognitive, and affective engagement with ChatGPT feedback. Their analysis identified difficulties in effective metacognitive regulation, despite the tool's capacity to engage learners. The small multiple-case design offers insight into processes rather than population-level estimates. It strengthens the rationale for explicit monitoring activities and for collecting evidence of how learners evaluate comments before revising.

Liu and Deris (2025) examined AI feedback literacy with 486 university students and found associations with feedback uptake. Their cross-sectional, self-reported measures limit causal inference and do not independently establish the quality of individual revision decisions. A related 2026 study by Yao used simulated writer agents rather than human learners to examine feedback scripts and a second-review mechanism. Its findings inform hypotheses about organised feedback processing, but cannot be counted as evidence of actual students' learning. These limitations make performance-based evaluation a necessary component of the proposed framework.

4.2 Five interdependent capabilities

For instructional purposes, AI feedback literacy is defined here as the situated capacity to elicit, interpret, verify, and use AI-generated comments through decisions that preserve intended meaning and support future writing. The synthesis organises this capacity into five domains, summarised in Table 1. These domains constitute a pedagogical map; they are not asserted to be empirically separable factors.

Table 1

Capabilities and observable evidence of AI feedback literacy

Capability	Learner action	Observable evidence
Task-aligned elicitation	Specify purpose, criteria, and feedback needs.	A focused request linked to a self-assessment goal.
Interpretation and prioritisation	Explain the comment and rank its importance.	A paraphrase, a located passage, and a reason for priority.
Evaluation and calibrated trust	Check the claim using relevant evidence.	An appropriate reference and a revised confidence judgement.
Agency, affect, and responsibility	Preserve meaning, manage reactions, and use assistance responsibly.	A reasoned disagreement or disclosure and a plan for resolving uncertainty.
Revision, reflection, and transfer	Revise deliberately and reuse the principle.	A justified edit and evidence of application in a new independent text.

Note. The domains are proposed pedagogical categories, not validated psychometric factors.

The first domain, task-aligned elicitation, involves deciding what feedback is needed and supplying relevant context. Requests identify genre, audience, assessment criteria, and the learner's own concern. The second, interpretation and prioritisation, involves paraphrasing a comment, locating the issue in the draft, and distinguishing a necessary correction from an optional stylistic preference. Both make the feedback manageable before the learner considers revision.

The third domain, evidence-based evaluation, concerns the justification for believing a comment. Verification varies with the issue: an authoritative grammar reference can test a grammatical claim, a genre exemplar can illuminate an organisational choice, and the original research publication can establish whether evidence has been represented accurately. Task-sensitive verification is central because no single checking strategy works equally well for all feedback.

The fourth domain connects agency with affect and ethical responsibility. Learners need permission to disagree, manage discouragement, and protect their intended stance. They also decide what information may be shared and acknowledge permitted assistance. The fifth domain concerns action over time: executing an informed revision, evaluating its effect, and applying the relevant principle in subsequent writing. The domains overlap deliberately because a revision decision can simultaneously involve language knowledge, confidence, ownership, and planning.

4.3 A recursive instructional cycle

The cycle begins with goal setting and self-assessment. Before requesting feedback, learners produce an initial draft and identify one rhetorical goal and one linguistic goal. For example, they may seek a clearer relationship between a claim and its evidence while also checking tense consistency. They predict the likely weakness and record what would count as improvement. This creates an initial judgement against which external feedback can be compared.

The second phase is selective elicitation. Learners request a small, prioritised set of comments connected to their goals. A useful request asks the tool to identify a relevant passage and explain the problem without replacing the whole text. The teacher initially supplies a prompt structure, then gradually removes that support. The purpose is to create a feedback encounter in which the learner's inquiry remains visible and the amount of information is compatible with available attention.

The third phase is verification and deliberation. Learners translate each selected comment into a claim that can be evaluated, locate supporting or contradictory evidence, and consider its likely effect on meaning. They record provisional confidence before checking and reconsider it afterwards. Confidence is a reflective judgement rather than a model-generated probability. When evidence remains insufficient, deferring the change and requesting human guidance is an appropriate outcome.

The fourth phase is justified revision. Learners accept a suggestion, adapt it, reject it, or defer it, then explain the decision. Acceptance requires showing why the change addresses the identified problem. Adaptation requires showing what was preserved and what was altered. Rejection requires a relevant reason rather than general distrust. Deferral identifies the unresolved issue and the next checking step. Learners then write their revision and compare the resulting passage with the original intention.

The fifth phase is reflection and transfer. Learners articulate a reusable principle, such as maintaining appropriate uncertainty when evidence is limited, and apply it in a new paragraph without reopening the AI conversation. A later independent task tests whether the principle can be used beyond the original text. The cycle is recursive: difficulty in independent application can lead to a revised goal, another exemplar discussion, or renewed instruction rather than merely another request for correction.

4.4 Decision quality and calibrated trust

The cycle treats trust as conditional on evidence and the specific task. A learner may reasonably rely on a verified agreement correction while questioning a proposed change to the strength of a research claim. This differs from expressing a stable overall attitude toward an AI tool. The relevant question is whether reliance is proportionate to the evidence available for the particular suggestion.

A high acceptance rate is therefore an ambiguous outcome. It can reflect accurate feedback and sound judgement, or uncritical compliance. Similarly, rejection can be justified or mistaken. Assessment must examine the reasoning and resulting text. Where several revisions are defensible, the teacher should recognise reasoned alternatives rather than impose a single preferred wording. The framework's central instructional claim is that explaining these choices creates opportunities for learning; whether those opportunities produce measurable transfer remains an empirical question.

5 Pedagogical implementation and assessment

5.1 A six-week teaching sequence

Table 2 proposes a six-week sequence embedded in an existing writing course. Each week uses one approximately 90-minute workshop alongside ordinary drafting and revision. The duration is an implementation suggestion, not a tested dosage. The sequence moves from teacher modelling to peer-supported decisions and increasingly independent application.

Table 2

Proposed six-week instructional progression

Week and focus	Core activity	Evidence for formative assessment
1 Standards and diagnosis	Compare exemplars; draft without AI; identify two goals.	Baseline writing and self-assessment.
2 Interpreting feedback	Judge source-blind comments; observe teacher verification.	Paraphrased comments and prioritisation reasons.
3 Focused elicitation	Request limited feedback using a scaffolded prompt.	Prompt, selected feedback, and relevance judgement.
4 Verification and decisions	Check mixed-quality advice; discuss accept, adapt, reject, or defer decisions.	Evidence-based decision records.
5 Revision and reflection	Revise their own draft; explain decisions to a peer; extract a principle.	Draft comparison and a short reflection.
6 Independent application	Write on a new topic without AI; explain use of the principle.	Independent text and transfer explanation.

Note. Suggested follow-up: an independent task two to four weeks later. The schedule is untested.

Early activities use short teacher-selected texts and pre-screened feedback containing accurate comments, misleading comments, and optional style changes. Learners judge these comments before learning their source. The teacher then models verification aloud, showing how a plausible suggestion can be accepted in one context and rejected in another. Later activities use learners' own drafts, with feedback requests limited to current learning goals.

Peer discussion supports the transition toward independent judgement. Each learner explains one consequential revision and one decision to reject or defer a comment. The partner checks the justification against the task and asks for evidence where needed. The teacher samples these exchanges and addresses recurring problems rather than approving every routine decision. This use of dialogue is consistent with the learner responsibility emphasised in Winstone et al.'s (2017) review of agentic engagement with feedback.

Support should vary with proficiency. Less proficient learners can work with short passages, selected comments, accessible reference materials, and first-language explanations of reasoning. More proficient learners can evaluate argument structure, disciplinary stance, and competing genre expectations. Teachers should separate difficulty in expressing an explanation in English from difficulty in making the judgement itself, while still checking whether the learner can implement the chosen revision in English.

5.2 An illustrative task and revision record

Consider a teacher-designed argumentative task about access to university libraries. A learner writes: "Longer opening hours may help students who work during the day." Feedback proposes

replacing “may help” with “will significantly improve academic achievement” to make the argument more persuasive. The proposed wording changes a cautious claim about access into a strong causal claim about achievement. Without additional evidence, the learner can justify rejecting it while considering a clearer explanation of the original argument.

In the same illustrative draft, “Students needs a quiet place” can be revised to “Students need a quiet place” after checking subject–verb agreement. These examples are constructed teaching materials, not observations from a classroom study. Table 3 records the contrasting decisions and a further example of adapting useful but underspecified feedback.

Table 3

Illustrative record of feedback decisions

Feedback claim	Checking evidence	Decision and rationale
Change “Students needs” to “Students need.”	The plural subject takes the base verb form in this present-tense clause.	Accept. The correction improves agreement without altering meaning.
Change “may help” to “will significantly improve academic achievement.”	The draft provides no evidence for the stronger causal claim or magnitude.	Reject. Retain appropriate uncertainty; improve the explanation of access instead.
“Add a statistic to support the point.”	No relevant statistic has been located and verified.	Adapt. Seek a genuine source or use a supported explanation; do not invent a number.

Note. All entries are constructed teaching examples, not empirical participant data. Classroom records should name the actual source consulted.

A feedback request for this activity might read: “Provide feedback on my paragraph for an EFL writer at B1 to B2 level. My goal is to connect my claim with my evidence. Use these criteria: relevance, clear reasoning, and grammatical accuracy. Identify no more than three important issues. For each, point to the relevant words and explain the concern in simple English. Separate necessary corrections from optional style suggestions. Do not rewrite the paragraph or invent evidence. If a claim needs checking, identify what I should verify.” This is an instructional template, not a guarantee of the tool’s compliance.

5.3 Performance-based assessment

Assessment combines a feedback-judgement task, a revision portfolio, and independent writing. In the judgement task, learners evaluate teacher-prepared comments with known strengths and weaknesses. The portfolio contains the original passage, the selected AI feedback, checking evidence, a decision record, and the learner’s revision. It need not include an entire conversation: a short excerpt and two or three consequential decisions can provide sufficient evidence while reducing unnecessary collection of personal information.

Four dimensions guide formative assessment: interpretation, verification, justified action, and transfer. Interpretation concerns whether the learner identifies the issue and its relation to the task. Verification concerns whether the selected evidence actually tests the comment. Justified action

concerns whether the decision and revision preserve meaning while addressing the problem. Transfer concerns whether the learner applies the principle in a new text and can explain the connection.

Each dimension can initially be described at three levels: emerging, supported, and independent. An emerging judgement rests mainly on agreement with the tool; a supported judgement uses relevant evidence with prompts; an independent judgement selects appropriate evidence and explains the decision without prompting. These descriptors are proposed teaching aids, not a validated measurement instrument. Teachers should discuss sample performances and moderate decisions before using them for consequential assessment.

Writing quality is assessed separately through criteria for task fulfilment, organisation, evidence, and language. A delayed task on a new but comparable topic, completed without AI access, provides evidence distinct from immediate revision gains. Brief oral explanation can help establish understanding when the written product alone is inconclusive. Satisfaction questionnaires can supplement these measures, but cannot replace observation of decisions and independent performance.

6 Discussion

6.1 Contribution and proposed learning mechanisms

The framework connects capacities commonly discussed at a conceptual level with observable points of decision in EFL writing. Its contribution lies in the alignment between five capabilities, a recursive cycle, classroom activities, and assessment evidence. It does not assume that prompt training, feedback literacy, and writing development are identical. Instead, it makes their proposed relationship available for examination through records of what learners ask, check, decide, and subsequently produce.

Three mechanisms explain why the design may be educationally useful. Self-assessment establishes a learner-generated diagnosis before external advice. Verification introduces evidence between a fluent recommendation and its adoption. Reflection requires converting a text-specific decision into a principle usable later. These are theoretical mechanisms derived through synthesis, not findings from the present article. Their plausibility depends on learners having enough language knowledge, suitable checking resources, and time to complete the relevant actions.

The framework also broadens the meaning of productive uptake. In instructional terms, a verified correction, a carefully adapted suggestion, and a justified rejection can each demonstrate literacy. This does not mean that every refusal represents learning. An incorrect rejection still requires attention. The key analytic unit is the relationship among the comment, the learner's evidence, the chosen action, and the resulting meaning, rather than acceptance considered in isolation.

6.2 Tensions and boundary conditions

Verification creates a competence problem: learners are asked to evaluate advice precisely where their knowledge may be incomplete. General instructions to "be critical" do not resolve this difficulty. Teachers need to build prerequisite knowledge through explicit explanation, exemplars, and constrained comparison tasks. A learner unable to distinguish a causal claim from a cautious association may need instruction on evidence and stance before independently evaluating AI comments about persuasiveness.

There is also a workload tension. A complete decision record for every minor edit could distract from writing and become a compliance exercise. The proposed solution is selective documentation: learners explain a few consequential choices, while routine corrections can be grouped. Teacher review

targets the adequacy of evidence and recurrent misunderstandings. As learners become more competent, scaffolds are reduced so that monitoring becomes useful to the writer rather than an additional form to complete.

Ownership requires attention to the effects of stylistic advice. A change can be linguistically acceptable while weakening a culturally meaningful example or replacing the learner's stance with a generic argument. Classroom discussion should therefore ask whether the revision serves the intended audience and meaning. Formality and conformity are not sufficient criteria for quality. A learner who retains a simple, precise sentence for a clear communicative reason may demonstrate stronger judgement than one who adopts more sophisticated wording without understanding it.

Equitable implementation requires access to the reasoning tasks, not necessarily identical personal subscriptions. Teachers can distribute the same saved feedback for evaluation, provide shared reference materials, and offer an equivalent non-AI pathway. Live use should follow institutional requirements for data handling and disclosure. Learners should remove identifying or confidential information from excerpts. These design choices address participation and responsibility within the framework; their adequacy requires review in the actual institutional context.

6.3 Empirical validation and research propositions

Three propositions specify a research agenda. First, learners receiving verification and justification instruction should make more accurate decisions about mixed-quality feedback than learners receiving prompt training alone. Second, decision quality should relate to later independent writing more strongly than raw acceptance rates. Third, scaffold reduction should support independent regulation when prerequisite language knowledge and checking resources are adequate. These propositions are hypotheses to test, with proficiency, prior AI experience, and genre knowledge treated as possible moderators.

A suitable study would compare the full framework with a prompt-training condition and an established teacher or peer feedback condition. Writing time, topics, criteria, and opportunities for revision should be comparable. Assignment should be random where feasible; an intact-class design needs to account for clustering and baseline differences. Sample size should follow a justified power analysis rather than an arbitrary classroom target. Relevant model versions, prompts, and interaction dates should be recorded because generated feedback is sensitive to its technological conditions.

Pre-intervention measures would establish writing proficiency and feedback judgement. Immediate measures would assess revision decisions and writing quality; a delayed assessment on a new topic would examine independent application. Human raters, blinded to condition where possible, should evaluate writing and decision records, with inter-rater agreement reported. Interviews or stimulated recall could clarify why learners accepted or rejected particular comments. Disagreement about rhetorical choices should be preserved analytically where more than one decision is defensible.

Evidence would remain incomplete if it showed only higher portfolio scores or favourable attitudes. The strongest support would combine improved judgement, appropriate revision, and delayed performance without AI assistance. Conversely, improved drafts accompanied by weaker independent performance would challenge the proposed mechanism and indicate that assistance had displaced important writing processes. Such a design would test the framework's educational claim while recognising that outcomes may differ across learners and writing contexts.

7 Conclusion

Developing AI feedback literacy requires instruction in the decisions connecting external advice with writing and learning. This article has proposed five capabilities and a recursive cycle through which EFL learners set goals, request focused feedback, test recommendations, justify revisions, and apply learning independently. The synthesis supports making feedback decisions observable while keeping evidence from older automated tools, generative systems, self-report studies, and simulations appropriately distinct.

The practical recommendations are to begin feedback work with self-assessment, teach verification through contrasting examples, preserve several legitimate revision choices, document a small number of consequential decisions, and assess subsequent writing without AI assistance. Teachers should adjust scaffolding to proficiency and progressively transfer responsibility for judgement to learners.

The framework offers a coherent basis for curriculum design and empirical investigation, but its effectiveness is not yet established. Its strongest claim is pedagogical: AI-supported revision becomes a more credible opportunity for learning when learners can explain what they changed, why they changed it, and how the underlying principle applies to a new writing task. Independent performance, rather than the appearance of a polished draft alone, must remain central to evaluating that opportunity.

References

1. Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. Assessment & Evaluation in Higher Education, 49(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321
2. Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. Assessment & Evaluation in Higher Education, 43(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354
3. Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0
4. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
5. Liu, K., & Deris, F. D. (2025). AI feedback literacy in higher education: Understanding, measuring, and predicting student feedback uptake. Assessment & Evaluation in Higher Education. Advance online publication. https://doi.org/10.1080/02602938.2025.2587924
6. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
7. Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. Smart Learning Environments, 11, Article 9. https://doi.org/10.1186/s40561-024-00295-9

8. Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. Assessment & Evaluation in Higher Education, 45(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955
9. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041
10. Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090
11. Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology, 8, Article 422. https://doi.org/10.3389/fpsyg.2017.00422
12. Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. Journal of Second Language Writing, 52, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816
13. Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. ReCALL, 36(2), 187–209. https://doi.org/10.1017/S0958344023000265
14. Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. Learning and Instruction, 91, Article 101894. https://doi.org/10.1016/j.learninstruc.2024.101894
15. Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. Higher Education, 76(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3
16. Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. Educational Psychologist, 52(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538
17. Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. Humanities and Social Sciences Communications, 11, Article 1086. https://doi.org/10.1057/s41599-024-03543-y
18. Yao, X.-C. (2026). Effects of feedback literacy scripts and a second-rater mechanism on EAP writing revision in generative AI-supported contexts. Frontiers in Psychology, 17, Article 1835593. https://doi.org/10.3389/fpsyg.2026.1835593
19. Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. Assessing Writing, 36, 90–102. https://doi.org/10.1016/j.asw.2018.02.004
20. Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2