

Indonesian Language Education and Applied Linguistics Reviews
Volume 2, Issue 2, 2026, 84-97
https://doi.org/10.66272/bgb6sr33

Available online at:
https://journal.mediaakademika.id/index.php/ileal/article/view/89

MEDIA AKADEMIKA PUBLISHER
ISSN 3110-5904

Who Controls the Revision When GenAI Gives Feedback? Evidence from Indonesian EFL Writers

Elfa Eriyani
Universitas Merangin
**Corresponding Author:* eriyani4366@gmail.com

ARTICLE INFO

Received:
29 August 2026

Revised:
27 October 2026

Accepted:
29 October 2026

Keywords:
generative AI, formative feedback, revision agency, EFL writing, academic integrity

ABSTRACT

This convergent mixed-methods multiple-case study examined how 16 Indonesian EFL undergraduates exercised revision agency while using generative-AI feedback across four connected writing tasks. The corpus comprised 64 final texts, 48.5 hours of screen recording, 342 prompt-response decision episodes, 16 interviews, stimulated recall, four focus-group reflections, and a 96-page analytic memo trail. Descriptive episode coding identified four patterns: selective scaffolding (38.6%), iterative co-composition (29.2%), surface repair (21.3%), and delegated composition (10.8%). Integrated qualitative analysis showed that high-agency use was marked by a bounded purpose, comparison of alternatives, transformation or rejection of output, verification, and the ability to defend the final text. High-proficiency students most often used selective scaffolding and rejected about 65% of outputs; intermediate students more often negotiated through iterative co-composition and rejected about 35%; low-intermediate students more often relied on surface repair or delegation and rejected about 10%. Frequency of AI use was therefore a weak indicator of learning or responsibility. Revision agency depended on what students did between receiving feedback and submitting text. The study contributes a process-based taxonomy and an evidence model for distinguishing assistance from substitution. It recommends bounded prompting, accept-modify-reject rationales, verification routines, teacher moderation, revision memos, and independent performance evidence.

How to Cite: Eriyani, E. (2026). Who Controls the Revision When GenAI Gives Feedback? Evidence from Indonesian EFL Writers. *Indonesian Language Education and Applied Linguistics Reviews, 2*(2), 84-97. https://doi.org/10.66272/bgb6sr33

Published by:
Media Akademika Publisher

1. INTRODUCTION

Generative artificial intelligence (GenAI) has altered the ecology of second-language writing. A student can now request explanations, alternative formulations, counterarguments, organisation advice, or complete passages at any point in composing. Unlike earlier automated writing evaluation systems, conversational models can respond to follow-up questions and generate extended text rather than merely flagging errors. These affordances are especially attractive in English-as-a-foreign-language (EFL) settings where students need repeated language support and teachers cannot provide immediate individual feedback on every draft. Research has consequently reported gains in accessibility, linguistic accuracy, revision opportunity, and confidence when AI-supported tools are incorporated into writing activity (Gayed et al., 2022; Kohnke et al., 2023; Meniado et al., 2024).

The same affordances complicate the meaning of feedback. Feedback traditionally provides information that learners must interpret and use. GenAI can instead perform part of the composing work itself. It may propose a transition, but it may also generate the paragraph that the transition connects; it may explain a weakness, but it may also replace the student's reasoning with fluent prose. The boundary between feedback and text production is therefore porous. Studies comparing ChatGPT and human commentary indicate that GenAI can offer rapid and

extensive feedback, particularly on language, while teacher involvement remains important for content, organisation, disciplinary expectations, and higher-order judgement (Guo & Wang, 2024; Han & Li, 2024; Steiss et al., 2024; Ding et al., 2025). The educational question is not simply whether the output is useful. It is whether the learner retains control over the intellectual and rhetorical decisions embodied in the final text.

1.1. From feedback provision to learner uptake

Contemporary feedback scholarship treats feedback as a process rather than a delivered message. Learners must recognise the purpose of feedback, make evaluative judgements, manage uncertainty, and take action (Carless & Boud, 2018; Molloy et al., 2020). Winstone et al. (2017) similarly locate agency in recipience processes through which learners seek, interpret, and implement information. This perspective matters for AI-supported writing because a technically strong response does not automatically become educationally productive. A suggestion may be ignored, copied, partially adapted, checked against a source, or used to generate a new idea. These different forms of uptake create different opportunities for learning even when the prompt and output appear similar.

Feedback agency is situated rather than purely personal (Nieminen et al., 2022). In GenAI-mediated writing, it depends on the learner's proficiency, task knowledge, assessment expectations, platform affordances, and available guidance. A strong rhetorical plan can support critical testing of alternatives, whereas limited language and genre knowledge can make equally fluent output difficult to evaluate. GenAI may therefore amplify resources while concealing gaps in understanding.

Second-language writing research distinguishes feedback exposure from behavioural, cognitive, and affective engagement (Fu & Liu, 2022). Students may accept a correction without understanding it or reject advice that conflicts with their intention. GenAI intensifies this problem because its responses are immediate, fluent, and apparently confident. Perceived source credibility can shape uptake even when credibility and accuracy diverge (Ruwe & Mayweg-Paus, 2024). Analysis must therefore examine the decisions connecting assistance with the final text.

1.2. GenAI-supported writing, authorship, and responsibility

Empirical studies offer a mixed picture. AI assistants can reduce lower-level linguistic burdens and free attention for ideas and organisation (Gayed et al., 2022). ChatGPT-supported feedback can improve mechanics, tone, grammar, and formatting, although content and organisational gains are less consistent without human mediation (Ding et al., 2025). EFL learners value speed, vocabulary support, paraphrasing, and conversational explanation (Meniado et al., 2024; Tsai et al., 2024). Yet revision studies show uneven engagement: some writers interrogate feedback and revise selectively, whereas others prioritise surface changes or depend on generated formulations (Koltovskaia et al., 2024). Reviews likewise identify concerns about transparency, relevance, over-reliance, and the balance between AI and human input (Lo et al., 2024).

These differences cannot be captured adequately through access statistics or self-reported frequency. A student may consult AI repeatedly while retaining a clear plan, rejecting most outputs, and using dialogue to sharpen judgement. Another student may make a single request and transfer a complete passage. Frequency would classify the first as the heavier user, even though the second episode transfers more consequential authorship. Product scores alone are also insufficient. A polished text may indicate improved learning, effective assistance, or concealed substitution. The interpretive problem is to identify what decisions remained with the student and what evidence shows that the student can explain and defend them.

Academic-integrity debates often frame GenAI through a binary of permitted assistance versus prohibited cheating. Clear rules remain necessary, but a binary classification can overlook the process through which permitted support becomes substitution. Ethical recommendations emphasise human responsibility, transparency, verification, and alignment with learning outcomes (Foltynek et al., 2023). Authorship discussions similarly connect ownership with substantive intellectual contribution and accountability rather than with the physical act of typing

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

(Bozkurt, 2024). These principles suggest that writing assessment should make the learner's decision process visible instead of relying only on detection or on declarations of tool use.

1.3. Revision agency as an observable process

This study defines revision agency as the learner's situated capacity to identify a writing need, set a bounded purpose for assistance, evaluate AI output against relevant criteria, decide whether to accept, modify, or reject it, verify consequential claims or language, and remain able to explain the final text. The definition combines feedback literacy with authorial responsibility. It does not assume that all AI involvement weakens agency, nor that all student-written language demonstrates strong agency. A learner can use AI extensively yet retain control through comparison and transformation. Conversely, a learner can make only minor visible edits while accepting a change in stance or meaning that they do not recognise.

Five observable indicators operationalise this definition. Purpose concerns whether the student can state what problem the interaction is intended to address. Evaluation concerns comparison of output with task criteria, sources, intended meaning, or alternatives. Transformation concerns whether the student reworks the output rather than transferring it unchanged. Verification concerns checking factual, linguistic, citation, or rhetorical accuracy. Explainability concerns the student's ability to account for the final choice during stimulated recall or interview. These indicators treat agency as evidence distributed across actions, artefacts, and explanations.

The patterns are modes of participation, not fixed student types. Writers can move between them as task difficulty, time pressure, and linguistic demand change. Proficiency provides evaluative resources but does not determine behaviour; analysis must therefore compare episodes and cases, including counterexamples.

1.4. Purpose, research questions, and contribution

Much of the emerging literature measures perceptions, compares AI and teacher feedback, or reports pre-post writing outcomes. These studies establish that GenAI can affect writing, but they reveal less about the moment-to-moment pathway through which generated feedback enters the text. Screen recordings, prompt-response logs, draft histories, interviews, and stimulated recall can show whether learners narrow a problem, compare alternatives, question a response, revise it substantially, or transfer it with little scrutiny. Combining episode counts with qualitative process evidence provides a more defensible account than either frequency or self-report alone.

The present study follows 16 Indonesian EFL undergraduates across four connected writing tasks and integrates descriptive coding of 342 AI-mediated decision episodes with cross-case qualitative analysis. It addresses three questions: (1) What patterns of GenAI-mediated feedback use and revision occur across the corpus? (2) How do students enact and explain acceptance, modification, rejection, verification, and ownership? and (3) How do dominant patterns vary across writing-proficiency groups? The study contributes an evidence-based taxonomy of four revision patterns, clarifies the relationship between proficiency and evaluative control, and proposes a process model that teachers can use to design instruction and assessment.

2. METHODS

2.1. Research design

A convergent mixed-methods multiple-case design was used. Each student constituted a case, while each GenAI-mediated revision decision was treated as an embedded unit of analysis. The quantitative strand described the distribution of coded episode patterns. The qualitative strand examined how those patterns were enacted and explained across screen recordings, drafts, interviews, stimulated recall, focus-group reflections, and analytic memos. Integration occurred at design, analysis, and interpretation through shared episode identifiers, cross-case matrices, and joint displays, consistent with the principle that mixed-methods integration should generate inferences that neither strand can provide alone (Fetters et al., 2013).

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

The design was descriptive and explanatory, not experimental. Because students used overlapping combinations of tools, the study neither compared platforms nor attributed outcomes to a particular model. The unit of interest was the decision process following AI-mediated assistance. Frequencies and percentages describe the corpus without population-level inference.

2.2. Setting and participants

The study was conducted in a six-week university EFL writing sequence in Indonesia. Sixteen undergraduates were selected through maximum-variation sampling to capture different levels of prior writing performance and different patterns of AI-tool engagement. Five were classified as high proficiency, seven as intermediate, and four as low-intermediate. Classification drew on cumulative grade-point average and an IELTS Writing equivalent used by the programme: high, GPA at least 3.70 and IELTS Writing equivalent 6.5-7.0; intermediate, GPA 3.00-3.69 and IELTS equivalent 5.5-6.0; and low-intermediate, GPA below 3.00 and IELTS equivalent 4.5-5.0. These categories supported within- and cross-case comparison and were not treated as immutable learner identities.

All participants used Grammarly Premium during the study. Twelve used ChatGPT-4o, nine used QuillBot, and five used Gemini; these counts overlap because most participants consulted more than one tool. Platform choice was not assigned by the researchers. Treating tool use as part of the authentic composing ecology increased ecological relevance but required analytical restraint: patterns were attributed to observable student decisions rather than to a platform name.

Table 1. Participant profile and tool engagement

Dimension	Category	n	Percentage
Writing proficiency	High	5	31.3%
Writing proficiency	Intermediate	7	43.8%
Writing proficiency	Low-intermediate	4	25.0%
Tool use	Grammarly Premium	16	100%
Tool use	ChatGPT-4o	12	75.0%
Tool use	QuillBot	9	56.3%
Tool use	Gemini	5	31.3%

2.3. Writing tasks and instructional conditions

Participants completed four connected tasks. The first was a problem-solution text that required a clearly bounded issue, evidence, and feasible recommendations. The second was an argumentative essay plan specifying claim, reasons, counterargument, and evidence. The third was a 900-1,100-word argumentative essay developed from the plan. The fourth required revision of the essay and a reflection on the reasons for consequential changes. Together the tasks made planning, drafting, feedback uptake, and reflective explanation visible across a sustained sequence rather than in a single submission.

AI use was permitted for brainstorming, explanation, language feedback, counterargument generation, and revision suggestions. Students remained responsible for evidence, citation accuracy, task compliance, and final submission. The instructor explained that AI output was provisional and that students might accept, adapt, or reject it. The study did not require a single standard prompt because the purpose was to observe authentic decision patterns. However, the class discussed the value of bounded requests, comparison of alternatives, and verification of claims. This common orientation reduced the risk that students would equate availability with reliability while preserving natural variation in practice.

2.4. Multimodal data corpus

The corpus combined product, process, and account data. Sixty-four final task texts represented four submissions for each participant. Draft histories were retained to trace the location and extent of revision. Screen recordings captured 48.5 hours of composing activity, with a mean of 3.03 hours per participant, including prompting, reading output, switching between resources, editing, and checking. Prompt-response logs contained 342 exchanges that could be

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

bounded as decision episodes. Sixteen semi-structured interviews explored perceived usefulness, trust, authorship, integrity, and responsibility. Stimulated recall used selected screen-recorded episodes to recover the student's purpose and reasoning close to the observed action. Four proficiency-banded focus-group reflections enabled participants to compare strategies and discuss shared dilemmas. Ninety-six pages of analytic memos documented coding decisions, counterexamples, and changes to the developing interpretation.

Table 2. Multimodal data corpus

Data source	Observed volume	Analytical purpose
Final task texts	64 texts (16 x 4)	Trace outcomes across repeated tasks
Draft histories	Four task histories per participant	Locate and compare revision
Screen recordings	48.5 hours; M = 3.03 hours per student	Observe prompting, checking, transfer, and revision sequence
Prompt-response logs	342 decision episodes	Describe and compare AI-use patterns
Interviews	16 sessions	Explore ownership, integrity, trust, and responsibility
Stimulated recall	Selected recorded episodes	Explain acceptance, rejection, and transformation
Focus-group reflections	4 proficiency-banded sessions	Test shared dilemmas and cross-case interpretation
Analytic memos	96 pages	Retain audit trail, counterexamples, and reflexive decisions

2.6. Qualitative analysis and mixed-methods integration

Reflexive thematic analysis interpreted meanings around purpose, trust, verification, transformation, responsibility, and voice (Braun & Clarke, 2021). Event summaries aligned screen actions with prompts, draft changes, and stimulated-recall accounts. Patterns were then compared within participants across four tasks and through cross-case matrices across proficiency groups. Focus-group reflections tested whether interpretations captured recognised dilemmas, while negative cases prevented deterministic labels.

Integration used weaving: frequencies showed how often patterns occurred, while process evidence explained their meaning. A joint display connected category, observed action, participant explanation, text consequence, and proficiency. Meta-inferences required support from two data types unless a contradiction was itself informative. Thus, a positive interview account did not establish high agency when recordings and drafts showed unchecked transfer.

2.7. Trustworthiness, reflexivity, and boundaries of interpretation

Trustworthiness procedures included multimodal triangulation, retained codebook versions, 96 pages of analytic memos, and documented decisions on ambiguous episodes. Negative-case review sought high-proficiency delegation and lower-proficiency selective control, testing the emerging interpretation. Cross-case matrices displayed recurrence and variation, while counts remained descriptive and tool choice was not treated causally.

Reflexivity addressed the normative assumptions attached to responsibility and authorship. Analysis neither equated typing with agency nor AI use with misconduct; it sought evidence of purpose, judgement, transformation, verification, and explainability. To protect identifying topics and personal information, the anonymous report uses cross-case descriptions rather than verbatim screen extracts. De-identified episode links remain available for confidential review subject to ethics restrictions.

2.8. Ethics

Institutional permission and informed consent were obtained before data collection. Participants consented separately to the collection of writing artefacts, prompt-response logs, screen recordings, interviews, stimulated recall, and focus-group reflections. Participation was separated from course grading. Identifiers were replaced with codes; personal information visible in recordings was redacted; and files were stored in access-controlled folders. Participants could

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

pause recording or withdraw material. The institutional approval identifier is withheld in the anonymous manuscript and should be supplied through the journal's confidential submission field.

3. RESULTS

3.1. Distribution of revision-agency patterns

The analysis of 342 GenAI-mediated decision episodes identified four distinct revision-agency patterns: selective scaffolding, iterative co-composition, surface repair, and delegated composition. Selective scaffolding was the most frequent pattern, representing 132 episodes (38.6%), followed by iterative co-composition with 100 episodes (29.2%). Surface repair accounted for 73 episodes (21.3%), while delegated composition was the least frequent pattern, with 37 episodes (10.8%). The percentages total 99.9% because of rounding. Overall, the distribution indicates that students more frequently retained or negotiated control over revision than transferred substantial responsibility to GenAI.

Figure 1 provides a visual comparison of the four patterns and highlights the relative dominance of selective scaffolding and iterative co-composition. Together, these two patterns accounted for more than two-thirds of all recorded episodes, suggesting that GenAI was most commonly used as a bounded source of assistance or as part of an iterative negotiation process rather than as a direct substitute for student writing. Surface repair occupied an intermediate position, whereas delegated composition represented a comparatively small proportion of the observed interactions.

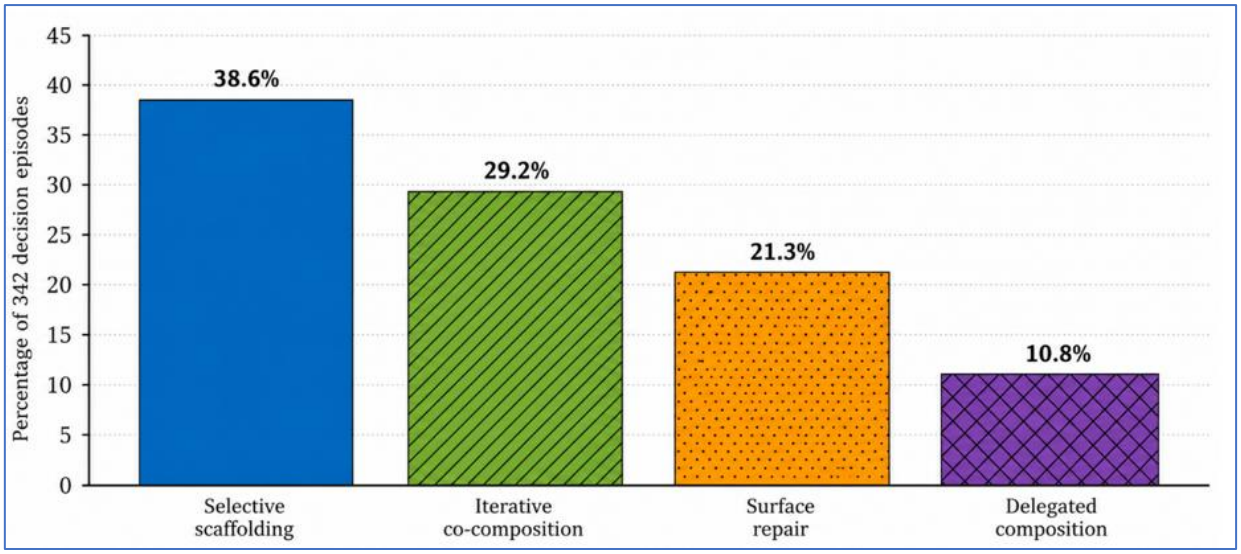


Figure 1. Distribution of four GenAI-mediated revision patterns across 342 decision episodes.

The distribution shown in Figure 1 should not, however, be interpreted solely in terms of frequency. Although delegated composition represented only 10.8% of the corpus, these episodes involved the greatest transfer of intellectual and rhetorical responsibility from the student to the AI system. Conversely, the higher frequency of selective scaffolding reflected instances in which students maintained clearer control over their goals, arguments, and final textual decisions. Thus, the educational significance of each pattern depended not only on how often it occurred but also on the degree of learner control evident within the revision process.

To clarify these differences, Table 3 presents the frequency and proportion of each revision pattern together with its corresponding interpretation of student agency. The classification was based on observable differences in how students defined the purpose of AI use, evaluated generated suggestions, transformed or rejected outputs, verified proposed changes, and retained responsibility for the final text. This framework therefore distinguishes qualitatively different forms of GenAI-supported revision rather than treating all AI-assisted writing as a single category.

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

Table 3. GenAI-mediated decision patterns

Pattern	n	%	Agency interpretation
Selective scaffolding	132	38.6	High: bounded assistance; student retains goal and argument
Iterative co-composition	100	29.2	Moderate-high: repeated negotiation and substantive transformation
Surface repair	73	21.3	Moderate: ideas remain student-owned but language control is uneven
Delegated composition	37	10.8	Low: AI performs consequential rhetorical work with limited verification

As Table 3 demonstrates, a simple assisted-versus-unassisted distinction would not adequately represent the observed revision processes because all four patterns involved GenAI while differing substantially in the distribution of decision-making authority. Selective scaffolding primarily expanded students' options without displacing their control, whereas iterative co-composition involved repeated interaction in which students continued to negotiate and reshape generated material. Surface repair preserved student ownership of the underlying ideas but sometimes reduced visible control over linguistic expression. Delegated composition represented the clearest reduction in revision agency because consequential rhetorical decisions were increasingly transferred to the AI system with limited evaluation, transformation, or verification by the student.

3.2. Selective scaffolding retained decision ownership

Selective scaffolding episodes typically began with a problem the student could already define. Requests were bounded: alternatives for a transition, possible objections to an existing claim, an explanation of a grammar problem, examples of how to clarify a sentence, or critique against stated task criteria. The screen record showed that students entered the interaction with a working paragraph or plan rather than asking the tool to establish the entire direction. This initial ownership constrained what the tool was asked to do.

Evaluation was visible after output appeared. Students compared alternatives, returned to the rubric, checked whether a suggestion matched the intended stance, or used the response to generate their own formulation. AI wording often did not survive into the final draft. Its function was diagnostic or catalytic: it made a problem visible, widened the choice set, or supplied a contrast against which the writer could decide. Rejection was therefore not a failure of the interaction. It was positive evidence that assistance remained provisional.

High-proficiency participants used this pattern most consistently and rejected approximately 65% of outputs. They noticed overstatement, generic examples, unnecessary formality, and mismatches with the essay's argument, while moving readily among AI output, sources, and their plan. Selective scaffolding was not exclusive to this group: lower-proficiency students also demonstrated it when tasks were familiar, prompts narrow, and comparison criteria explicit.

The defining feature was not minimal tool use but preserved decision ownership. Some selective-scaffolding episodes involved several requests. Agency remained high because the student controlled the question, treated outputs as alternatives, and could explain the final choice. This finding separates interaction volume from dependency: repeated use can support learning when it is organised around bounded purposes and evaluative action.

3.3. Iterative co-composition produced negotiated authorship

Iterative co-composition involved multiple cycles of prompting, response, reaction, and revision. Students asked for alternatives, questioned a generic claim, narrowed a request, combined parts of several suggestions, or returned an edited passage for further critique. Unlike selective scaffolding, the direction sometimes developed through dialogue rather than being fully specified in advance. AI contributed more visible wording and organisation, yet the first response was rarely treated as final.

Intermediate participants displayed this pattern most often and rejected about 35% of outputs. They generally possessed enough language and task knowledge to recognise an

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

unsuitable answer, but not always enough confidence to solve the problem independently. Dialogue functioned as a bridge. Students externalised a partially formed idea, inspected the response, and progressively refined both the prompt and the text. The resulting authorship was distributed, but it remained explainable when the learner could identify what had changed and why.

Negotiated authorship was strongest when students altered conceptual relationships rather than merely combining fluent fragments. Process evidence showed attention to order of reasons, relevance of examples, strength of claims, and relationship between counterargument and rebuttal. Agency weakened when iteration became a search for increasingly polished replacement language without an explicit criterion. Repeated prompting is therefore not automatically reflective. It becomes agentic when each turn embodies a judgement that changes the next request or the developing text.

This pattern occupies an important middle position. It challenges the assumption that legitimate authorship requires solitary production, while also rejecting the view that any conversational exchange establishes collaboration. The decisive evidence lies in the student's evaluative contribution: articulating criteria, diagnosing limitations, transforming output, and remaining able to account for the final organisation and language.

3.4. Surface repair improved readability but could displace voice

Surface repair focused on grammar, punctuation, vocabulary, phrasing, sentence structure, and paraphrase. It usually preserved student-generated ideas and therefore appeared less threatening to authorship than delegated composition. The episodes often produced immediate improvements in readability. For EFL writers, this support also reduced the cognitive burden associated with monitoring local form while developing an argument.

The process record revealed a subtler ownership problem. Students sometimes accepted academic-sounding phrasing without checking certainty, evaluation, stance, or disciplinary fit. A modal could become a stronger claim, or a cautious interpretation a universal statement. Such local changes were rhetorically consequential; surface accuracy did not guarantee fidelity to the writer's meaning.

Agency strengthened when students requested explanations, compared revisions with their intention, or accepted only the corrected element. It weakened when fluency became the main criterion for sentence replacement. Correction and control are therefore distinct: a student may own an idea while losing control over its rhetorical positioning. Pedagogy should treat voice as observable choices, not a vague personal quality.

Low-intermediate participants relied more heavily on surface repair because it addressed genuine linguistic needs. This reliance should not be interpreted as laziness. Their difficulty was evaluative asymmetry: the tool could produce language more advanced than they could reliably judge. Pedagogical support should consequently provide accessible comparison criteria and explanations, not simply prohibit assistance that may enable participation.

3.5. Delegated composition reduced accountability

Delegated composition involved complete paragraphs, substantial claim development, examples, or organisational decisions generated from broad requests and transferred with limited scrutiny. It was the least frequent pattern, with 37 episodes (10.8%), but it presented the clearest risk to authorship and learning. The concern was not merely textual overlap. The process evidence showed that the learner had relinquished consequential reasoning or rhetorical work that the task was intended to assess.

During stimulated recall, low-agency episodes were marked by difficulty explaining why a claim, example, or organisational choice appeared in the final draft. Cosmetic editing did not restore ownership when the student could not reconstruct the reasoning. Verification was also limited. Fluent claims were sometimes treated as ready for use without checking evidence, source fit, or alignment with the original plan. The final text could therefore appear stronger than the student's independent capacity.

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

Low-intermediate participants were most represented in this pattern and rejected about 10% of outputs. Linguistic need, time pressure, and polished English made direct transfer attractive. Negative cases showed that proficiency did not determine delegation: more proficient students delegated under pressure, while less proficient students retained agency with bounded prompts and explicit criteria. Delegation was therefore an episode-level condition shaped by resources, task demand, and instructional design.

The accountability problem was visible when three links were absent: the learner could not connect the output to a self-defined need, could not show how it had been evaluated, and could not defend its inclusion. This evidence-based definition is more useful than judging responsibility from the amount of AI-generated wording alone. It distinguishes a heavily revised AI suggestion that the learner can justify from a short but consequential claim accepted without understanding.

3.6. Cross-case variation by proficiency

Proficiency shaped the resources available for agency. High-proficiency students had greater capacity to diagnose rhetorical mismatch, compare alternatives, detect generic output, and protect personal voice. Intermediate students often used dialogue to compensate for incomplete certainty and developed arguments through negotiated cycles. Low-intermediate students benefited from language support but faced the greatest difficulty evaluating fluency that exceeded their productive competence. Table 4 integrates the dominant pattern, approximate rejection rate, and process interpretation for each group.

Table 4. Cross-case integration by proficiency

Group	Dominant pattern	Output rejection	Integrated interpretation
High (n=5)	Selective scaffolding	About 65%	AI broadened options; students protected argument, verification, and personal voice
Intermediate (n=7)	Iterative co-composition	About 35%	Students negotiated organisation and language; ownership was shared but usually explainable
Low-intermediate (n=4)	Surface repair/ delegation	About 10%	Linguistic need increased direct uptake and reduced capacity to evaluate output

These tendencies are descriptive, not hierarchical. Task familiarity, prompt scope, time, confidence, and teacher criteria produced substantial within-group variation. Proficiency mattered most during evaluation: all groups could obtain output, but they differed in recognising inaccuracy, rhetorical mismatch, or altered meaning. The result supports differentiated scaffolding rather than a uniform AI policy.

3.7. Integrated process model of revision agency

Across categories, revision agency was preserved through a five-part decision sequence: define a writing need, request bounded support, evaluate and compare output, modify or reject it selectively, and verify and defend the final text. The sequence is analytical rather than rigidly linear; students could revisit an earlier need or compare several outputs. Its value lies in identifying the evidence that should remain visible when AI participates in writing. Figure 2 presents the model in a deliberately simple form.

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

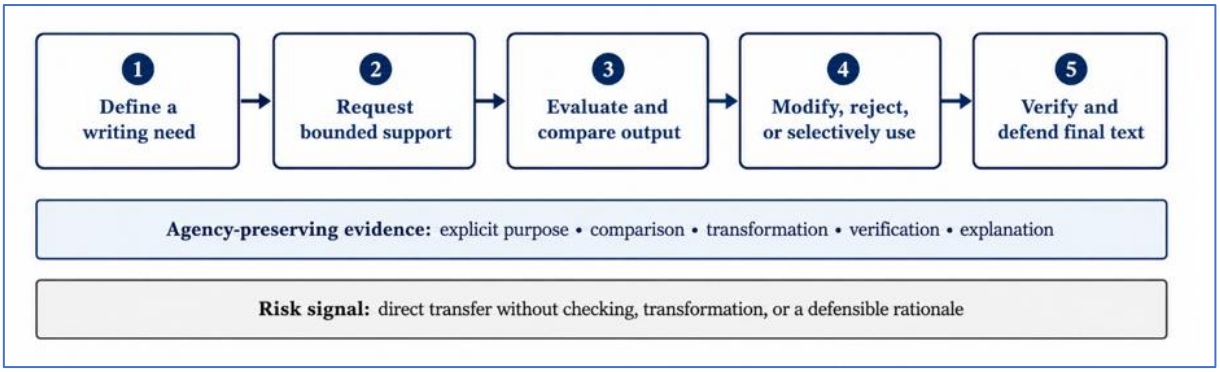


Figure 2. Process model of revision agency in GenAI-mediated feedback.

The model also locates the principal risk. GenAI becomes substitutive when the pathway is shortened from output to direct transfer and the evaluative stages disappear. This short circuit can occur in any proficiency group, although limited linguistic resources increase vulnerability. Conversely, transformation and rejection demonstrate that the learner is not merely consuming output. Verification and explanation reconnect the final text to responsibility. These stages provide a practical basis for revision memos, oral defence, and process-oriented assessment.

Table 5. Pattern-sensitive pedagogical responses

Observed pattern	Primary learning need	Recommended response
Selective scaffolding	Refine evaluative judgement	Require comparison criteria and retain brief rationale for acceptance or rejection
Iterative co-composition	Keep dialogue purposeful and explainable	Use checkpoints for claim, evidence, structure, and final student justification
Surface repair	Protect meaning, stance, and voice	Compare original and revised sentence; request explanation before replacement
Delegated composition	Restore intellectual participation	Narrow the task, stage support, verify evidence, and pair with independent performance

4. DISCUSSIONS

4.1. AI-use frequency is a weak proxy for agency

The central contribution is a shift from counting tool use to examining the quality of the decisions that follow it. Most episodes involved selective scaffolding or iterative co-composition, but a smaller delegated category carried disproportionate authorship risk. This pattern explains why neither frequency nor a simple declaration of AI use can establish responsibility. A frequent user may exercise strong agency through comparison and rejection; an infrequent user may transfer a consequential paragraph. The relevant unit for pedagogy and assessment is the decision episode linking a writing need, an external response, and the final text.

This process account extends feedback-literacy theory. Carless and Boud (2018) emphasise judgement and action, while Molloy et al. (2020) describe feedback literacy as a set of learner capabilities. GenAI adds a distinctive challenge because the feedback source can generate the object being assessed. Revision agency therefore requires not only understanding comments but also regulating the division of labour between writer and tool. Purpose, evaluation, transformation, verification, and explainability make that division observable.

The findings also support a sociomaterial view of agency (Nieminen et al., 2022). Agency depended on proficiency, task knowledge, time, criteria, platform fluency, and teacher mediation. The same fluent response could scaffold one learner and substitute for another, and the same student could move between patterns. Policy should therefore create conditions for judgement rather than label users as responsible or dependent.

4.2. Proficiency is a conditional evaluative resource

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

Proficiency differences were most visible in evaluation and rejection. High-proficiency students could detect misalignment and preserve rhetorical direction; intermediate students often built judgement through dialogue; low-intermediate students faced an asymmetry between the quality of generated English and their ability to appraise it. This result is consistent with research showing that engagement with automated feedback varies across learners and that lower-order assistance is easier to understand and enact than feedback on content or organisation (Fu & Liu, 2022; Ding et al., 2025).

The implication is not that lower-proficiency learners should be denied AI support. Such a policy could remove a meaningful participation scaffold while leaving underlying evaluative needs unaddressed. Instead, support should be differentiated. Bounded prompt frames can reduce the scope of generated intervention. Side-by-side comparison can make changes visible. Plain-language explanations can support noticing. Teacher or peer checkpoints can concentrate on stance, evidence, and organisation. Movement from AI-assisted to independent performance can then be staged rather than assumed.

4.3. Designing pedagogy around visible decisions

The process model suggests a practical revision routine. Before consulting AI, students state the writing problem and the criterion that will guide evaluation. After receiving output, they record whether they accept, modify, or reject it. For consequential changes, they identify what was checked and explain how the final version serves the argument. This routine converts invisible interaction into assessable evidence without requiring surveillance of every keystroke.

A concise revision memo can record the intended problem, AI contribution, accept-modify-reject decision, verification evidence, and final rationale. Its purpose is to prompt judgement and support targeted discussion, not reward elaborate disclosure. Making rejection visible treats it as learning, while comparing plausible outputs prevents the first fluent response from becoming the default.

Teacher feedback remains important. AI can support repeated language noticing and low-stakes experimentation, while teachers concentrate on argument, evidence, genre, and disciplinary expectations. This allocation accords with evidence that GenAI often performs best on accessible, lower-order concerns and human mediation adds value for content and organisation (Guo & Wang, 2024; Han & Li, 2024; Ding et al., 2025).

Table 6 translates each pattern into a differentiated response: refine selective scaffolding, keep co-composition evaluative, protect meaning during surface repair, and pair delegation risks with staged support and independent evidence.

4.4. Integrity and assessment should focus on accountability

Academic-integrity policy is necessary, particularly for disclosure, privacy, citation, and prohibited task conditions (Cotton et al., 2024; Foltynek et al., 2023). However, detection-centred approaches are poorly aligned with the process revealed here. They attempt to infer assistance from the final product while the educationally important evidence lies in how the student made and defended decisions. Accountability is better supported by triangulating product with process evidence: revision history, a short rationale, source checks, oral explanation, or a related independent task.

Independent performance evidence is crucial because assisted quality and learning are not synonymous. A strong AI-supported essay may conceal reasoning the student cannot reproduce. Pairing coursework with an in-class plan, brief oral defence, annotated revision, or delayed unaided sample supports valid inference rather than punishment. It also prevents polished output from obscuring lower-proficiency learners' progress and remaining needs.

Authorship should consequently be treated as accountable intellectual participation. Human ownership is strongest when the learner defines the problem, determines criteria, transforms or rejects suggestions, verifies claims, and can explain the final text. This formulation moves beyond the false choice between fully independent writing and illegitimate machine production. It recognises responsible collaboration while retaining the principle that the student must remain answerable for submitted work.

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

4.5. Methodological and theoretical contribution

Methodologically, linking logs, recordings, draft histories, and stimulated recall at episode level made acceptance, transformation, rejection, and checking observable. This process triangulation counterbalanced the limits of self-report and product analysis. Descriptive counts indicated the relative presence of qualitative pathways without implying population representation.

Theoretically, the taxonomy replaces a use-versus-non-use binary with four qualitatively distinct relations: bounded assistance, negotiated authorship, redistributed linguistic control, and transferred consequential work. Classification depends on the learner's evaluative relation to output, not generated-text quantity alone, providing more precise terms for research, pedagogy, and policy.

4.6. Limitations and future research

Limitations include the small, single-program sample, six-week duration, and overlapping tool use; percentages describe this corpus only. Proficiency groups combined GPA with an IELTS Writing equivalent rather than independent psychometric classifications. Recording may have influenced behaviour, unrecorded use remained invisible, and stimulated recall was retrospective. Omitting verbatim screen excerpts protected participants but reduced direct access to raw interaction.

Future research should test the taxonomy across institutions, languages, genres, and assessment conditions. Longitudinal and experimental studies could examine whether revision memos, comparison routines, bounded prompting, or hybrid feedback improve supported performance and delayed independent transfer. Larger process datasets could model transitions among patterns and interactions between proficiency, task demand, time pressure, and AI literacy. Transfer of verification routines across changing platforms also warrants study.

5. CONCLUSION

GenAI-mediated feedback becomes educationally meaningful when students retain the capacity to define purposes, evaluate suggestions, transform or reject output, verify consequential claims and language, and defend the final text. Across 342 decision episodes, selective scaffolding and iterative co-composition were more common than delegation, but proficiency-related differences revealed unequal capacity to judge fluent output. The study therefore rejects AI-use frequency as a sufficient indicator of learning, integrity, or dependence.

The practical response is to design for visible revision agency. Bounded prompts, comparison of alternatives, accept-modify-reject rationales, verification routines, teacher moderation, and independent performance evidence can preserve assistance without allowing it to become invisible substitution. The proposed taxonomy and five-stage process model provide a basis for research and assessment that recognises responsible human-AI collaboration while keeping students accountable for the intellectual and rhetorical work submitted in their name.

6. REFERENCES

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Bozkurt, A. (2024). GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis, 16*(1), 1–10. https://doi.org/10.55982/openpraxis.16.1.654

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Ding, L., Zou, D., & Kohnke, L. (2025). ChatGPT as an automated writing evaluation tool: How students perceive it and how it affects their writing. *Education and Information Technologies, 30*, 26777–26799. https://doi.org/10.1007/s10639-025-13775-3

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—Principles and practices. *Health Services Research, 48*(6, Pt. 2), 2134–2156. https://doi.org/10.1111/1475-6773.12117

Foltynek, T., Bjelobaba, S., Glendinning, I., Khan, Z. R., Santos, R., Pavletic, P., & Kravjar, J. (2023). ENAI recommendations on the ethical use of artificial intelligence in education. *International Journal for Educational Integrity, 19*, Article 12. https://doi.org/10.1007/s40979-023-00133-4

Fu, H., & Liu, X. (2022). EFL learner engagement in automatic written evaluation. *Frontiers in Psychology, 13*, Article 871707. https://doi.org/10.3389/fpsyg.2022.871707

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence, 3*, Article 100055. https://doi.org/10.1016/j.caeai.2022.100055

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 126*, Article 103502. https://doi.org/10.1016/j.system.2024.103502

Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing, 30*, 31–44. https://doi.org/10.1016/j.jslw.2015.08.002

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing, 65*, Article 101130. https://doi.org/10.1016/j.jslw.2024.101130

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11*, Article 50. https://doi.org/10.1186/s40561-024-00342-5

Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence, 7*, Article 100313. https://doi.org/10.1016/j.caeai.2024.100313

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

Nieminen, J. H., Tai, J., Boud, D., & Henderson, M. (2022). Student agency in feedback: Beyond the individual. *Assessment & Evaluation in Higher Education, 47*(1), 95–108. https://doi.org/10.1080/02602938.2021.1887080

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.

guidelines. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/16094069198992 20

Ruwe, T., & Mayweg-Paus, E. (2024). Embracing LLM feedback: The role of feedback providers and provider information for feedback effectiveness. *Frontiers in Education, 9*, Article 1461362. https://doi.org/10.3389/feduc.2024.1461362

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

Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies, 29*, 22427–22445. https://doi.org/10.1007/s10639-024-12722-y

Wiboolyasarin, W., Wiboolyasarin, K., & Suwanwihok, K. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. *Computers and Education: Artificial Intelligence, 6*, Article 100228. https://doi.org/10.1016/j.caeai.2024.100228

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

Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*. Advance online publication. https://doi.org/10.1080/17501229.2025.2503890

Copyright (c) 2026 Eriyani
This work is licensed under a Creative Commons Attribution 4.0 International License.