Indonesian Language Education and Applied Linguistics Reviews
Volume 2, Issue 1, 2026, 33-45
https://doi.org/10.66272/fqd0qm51

MEDIA AKADEMIKA PUBLISHER
ISSN 3110-5904

Available online at:
https://journal.mediaakademika.id/index.php/ileal/article/view/79

Beyond Convenience: How EFL Students Negotiate Generative AI Support, Authorship, and Academic Integrity in English Writing

Kurnia Saputri
[1]*Universitas Muhammadiyah Palembang
**Corresponding Author:* kurniasaputri93@gmail.com

ARTICLE INFO

Received:
24 April 2026

Revised:
26 May 2026

Accepted:
15 June 2026

Keywords:
generative AI, EFL writing, learner agency, authorship, academic integrity, qualitative case study

ABSTRACT

Generative artificial intelligence (GenAI) has become embedded in university writing practices, yet perception-based research cannot show what learners actually do when AI suggestions enter the composing process. This qualitative multiple-case study examined how 16 EFL undergraduates used, evaluated, accepted, rejected, and transformed GenAI assistance across four English-writing tasks completed over six weeks, with particular attention to learner agency, authorship, and academic integrity. Participants represented high-proficiency (n = 5), intermediate-proficiency (n = 7), and low-intermediate-proficiency (n = 4) writing profiles. The multimodal corpus comprised 64 task-specific final texts, 48.5 hours of screen recordings, and 342 recorded prompt-response decision episodes, complemented by interviews and stimulated-recall accounts. Episode-level coding identified four patterns: selective scaffolding (132 episodes; 38.6%), iterative co-composition (100; 29.2%), surface repair (73; 21.3%), and delegated composition (37; 10.8%). The first two patterns showed the strongest learner agency because students bounded AI use, modified prompts, challenged inaccurate outputs, and reworked generated material. Delegated composition represented the clearest loss of authorship, with whole AI-generated paragraphs copied with little verification or substantive revision. A stimulated-recall episode illustrates how ownership was preserved through rejection and restructuring of AI output. The findings show that responsible GenAI use is better understood as a sequence of evaluative decisions than as a binary distinction between AI use and non-use. The study proposes an AI-mediated writing agency framework centred on purpose, verification, transformation, disclosure, and responsibility.

How to Cite: Saputri, K. (2025). Beyond Convenience: How EFL Students Negotiate Generative AI Support, Authorship, and Academic Integrity in English Writing. *Indonesian Language Education and Applied Linguistics Reviews, 2*(1), 33-45. https://doi.org/10.66272/fqd0qm51

Published by:
Media Akademika Publisher

1. INTRODUCTION

Writing in an additional language has always required learners to coordinate multiple demands at once: generating ideas, organizing arguments, selecting appropriate lexical and grammatical resources, anticipating readers, integrating sources, and revising texts into increasingly purposeful forms. For EFL university students, these demands are intensified by the need to make disciplinary and rhetorical decisions through a language in which they may still be developing confidence. Digital writing tools have long offered partial support, but the emergence of large language model-based generative artificial intelligence (GenAI) has changed the scale and character of that support. A student can now ask a system to brainstorm a position, produce an outline, rewrite a paragraph, explain a grammar problem, generate counterarguments, imitate a genre, summarize a source, or produce an entire draft within seconds.

The literature review from which this study is developed shows a consistent pattern. EFL learners generally value AI-assisted writing because it can reduce linguistic burden, provide immediate feedback, support idea generation, expand vocabulary, assist revision, and increase

opportunities for independent learning. Earlier automated writing evaluation research similarly demonstrated that learners do not always passively accept machine feedback: students may inspect, reject, or selectively implement suggestions depending on task goals and perceived accuracy (Fu & Liu, 2022; Hoang, 2022; Shang, 2022; Wang & Han, 2022). Work on AI-based writing assistants has also suggested that automated support can help learners redistribute attention from lower-level linguistic problems toward higher-order composing decisions (Gayed et al., 2022). These benefits help explain why GenAI has become attractive in EFL contexts where students often experience unequal access to individualized writing feedback.

Yet the same affordances create an authorship problem that is not adequately captured by asking whether students have positive or negative perceptions of AI. When a learner asks a model to generate ideas, rewrites its output, combines machine text with personal language, accepts a suggested sentence without checking it, or delegates an entire section, the central issue is no longer simple technology acceptance. The issue becomes one of agency: who determines the rhetorical direction of the text, who evaluates the truth and appropriateness of suggestions, who makes revision decisions, and who can account for the final product? Academic integrity concerns therefore cannot be reduced to a binary distinction between "AI use" and "no AI use." The ethical meaning of AI assistance depends on the function it performs, the degree of human judgement involved, the transparency of its use, and the extent to which the learner remains intellectually responsible for the submitted text.

This distinction is important because current institutional responses often concentrate on detection, prohibition, or broad declarations that students must "use AI responsibly." Such responses provide limited pedagogical guidance for the many ambiguous decisions that occur inside an actual writing episode. A student who asks for three possible thesis statements and then constructs a new one through comparison is engaging with AI differently from a student who pastes a complete prompt and submits the generated response with minimal modification. A student who uses AI to identify unclear sentences and independently rewrites them is also participating differently from a student who automatically accepts every correction. These differences are process differences, but most perception surveys cannot capture them.

A stronger empirical approach must therefore follow students across the chain of composing activity. It should connect prompts to outputs, outputs to student decisions, decisions to textual revisions, and revisions to students' retrospective explanations of ownership and responsibility. Such a design treats AI use not as a single variable but as a sequence of situated decisions. It also allows researchers to examine contradictions between what students say and what they do. Learners may describe themselves as cautious users while relying heavily on generated phrasing; conversely, students who report frequent AI use may demonstrate strong critical control by verifying, recombining, and rejecting machine suggestions.

The Indonesian EFL university context makes this inquiry particularly relevant. Students increasingly encounter GenAI across writing-intensive coursework, yet classroom expectations, lecturer policies, and students' AI literacy may differ substantially. The original narrative review identified the need for more contextual research in developing-country settings and noted concerns about over-reliance, originality, inaccurate information, and academic integrity. Rather than repeating a perception study, the present research design addresses that gap by investigating AI-mediated writing as observable practice.

Accordingly, this study asks four questions: (1) How do EFL university students integrate GenAI across idea generation, drafting, language support, and revision? (2) What evaluative strategies do students use when deciding whether to accept, reject, or transform AI-generated assistance? (3) How do students understand authorship, ownership, disclosure, and academic integrity in relation to their AI-supported texts? and (4) What recurring patterns of learner agency emerge across AI-mediated writing episodes? By answering these questions through process data rather than attitudes alone, the study aims to contribute a more precise account of what responsible GenAI use can look like in EFL writing pedagogy.

2. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

2.1. From automated feedback to generative writing partnership

Technology-mediated writing support predates contemporary GenAI. Automated corrective feedback, automated writing evaluation, paraphrasing systems, and grammar-checking tools have long provided learners with suggestions that influence revision. Research in these environments established an important principle: feedback does not improve writing simply because it is available. Its pedagogical value depends on learner engagement. Fu and Liu (2022), for example, foregrounded EFL learner engagement in automatic written evaluation, while Hoang (2022) showed that learners' responses to automated corrective feedback vary according to perceived precision and usefulness. Shang (2022) likewise connected automated and peer feedback with EFL writing performance, and Wang and Han (2022) examined cognitive and psychological dimensions of automated versus teacher feedback. Together, such studies position the learner as an interpreter of feedback rather than a passive endpoint of a technological system.

Generative AI substantially expands this relationship. Traditional automated feedback typically reacts to existing text; GenAI can participate before text exists. It can propose topics, outline arguments, simulate readers, generate examples, reformulate claims, or produce complete passages. This means that AI can operate at both lower-order and higher-order levels of composing. The distinction matters because assistance with grammar correction does not raise the same authorship questions as assistance that determines the conceptual architecture of an essay. Gayed et al. (2022) demonstrated the potential of an AI-based writing assistant to support English language learners in overcoming cognitive barriers, while Woo and colleagues' work on natural language generation and EFL composing drew attention to how learners use generated suggestions during idea development. The educational challenge is therefore not to classify all AI assistance as equivalent, but to identify the role AI occupies in different phases of writing.

The original review also emphasized perceived benefits for autonomy. Immediate access allows students to seek assistance without waiting for a teacher, potentially supporting self-directed learning. However, autonomy should not be confused with independence from human instruction or with unrestricted tool use. Meaningful autonomy requires learners to make informed choices, monitor their goals, evaluate external feedback, and take responsibility for outcomes. AI may strengthen autonomy when it expands the learner's capacity to diagnose problems and revise strategically; it may weaken autonomy when learners outsource judgement itself.

2.2. Learner agency as decision-making, not mere tool use

Agency provides a useful lens for distinguishing these possibilities. From a social-cognitive perspective, agency involves intentionality, forethought, self-regulation, and reflective evaluation rather than simple action (Bandura, 2001). Applied to AI-assisted writing, agency concerns whether students remain active designers of their composing process. An agentic writer can use substantial technological assistance while still determining goals, evaluating alternatives, and accepting responsibility. By contrast, a learner may appear "independent" because no teacher is involved but exercise little agency if machine output is accepted without interrogation.

For this study, AI-mediated writing agency is operationalized through five observable dimensions. Purpose refers to whether the learner can articulate why AI is being used at a particular moment. Verification concerns checking factual accuracy, source claims, linguistic appropriateness, and task alignment. Transformation refers to the degree to which generated material is reworked, integrated, personalized, or rejected rather than copied. Disclosure concerns the learner's awareness of course expectations and willingness to represent AI involvement transparently where required. Responsibility concerns the learner's capacity to explain and defend the final text as their own intellectual product. These dimensions turn an abstract idea of "responsible AI" into practices that can be observed in writing traces and discussed in interviews.

This framework also prevents a simplistic assumption that frequent AI use automatically indicates weak agency. A student may generate many prompts because they are iteratively testing ideas and comparing alternatives. Conversely, a student may use AI only once but delegate an entire essay. Frequency is therefore an insufficient measure. The unit of analysis should be the

decision episode: a bounded sequence in which a learner has a writing need, seeks AI assistance, receives output, evaluates it, and makes a textual or strategic decision.

2.3. Authorship, ownership, and academic integrity

Academic integrity debates surrounding AI often begin with originality, but authorship is broader than textual novelty. A writer is conventionally expected to be accountable for claims, evidence, organization, and wording presented under their name. GenAI complicates this expectation because machine-generated language can be fluent, plausible, and difficult to distinguish from human prose, while its epistemic reliability is uneven. Roe and Perkins (2022), writing about automated paraphrasing, already highlighted how technologies that transform text challenge traditional plagiarism and integrity practices. GenAI intensifies that challenge by generating content rather than merely rephrasing source text.

A process perspective helps separate different ethical risks. One risk is misrepresentation: presenting externally generated intellectual work as one's own when course rules prohibit or require disclosure of such use. A second is epistemic negligence: accepting claims or references without verification. A third is developmental displacement: using AI in ways that bypass the very cognitive practice an assignment is designed to develop. A fourth is dependency: gradually transferring planning, language choice, and revision judgement to a system until the learner becomes less able to perform those functions independently. These risks may overlap, but they are not identical and require different pedagogical responses.

The central question for writing pedagogy is therefore how to preserve meaningful authorship in a context where assistance is increasingly normal. A useful answer is not an arbitrary percentage of human versus AI-generated words. Authorship is better understood through decision ownership. Students should be able to show how a text developed, explain why major claims and structures were selected, verify information, revise language in relation to communicative purposes, and account for any AI assistance. This places process evidence, reflective explanation, and evaluative judgement at the center of integrity.

2.4. Research gap and contribution

The literature contains growing evidence that students appreciate AI for efficiency, linguistic support, and feedback, while also expressing concern about dependence, inaccuracy, and academic dishonesty. What remains less visible is the micro-process connecting those perceptions to actual writing behavior. Studies that rely on post-use questionnaires cannot determine whether students critically evaluate outputs. Interview-only designs may capture retrospective accounts but miss fleeting decisions that students forget or rationalize. Product-only comparisons can show changes in text quality but not who or what generated those changes.

The present study addresses this gap through methodological triangulation across screen activity, AI logs, drafts, revisions, interviews, and stimulated recall. Its proposed contribution is twofold. Empirically, it identifies patterns in how EFL students distribute writing work between themselves and GenAI. Conceptually, it proposes a five-dimension framework for AI-mediated writing agency—purpose, verification, transformation, disclosure, and responsibility—that can support both research coding and classroom instruction.

3. METHODS

3.1. Research design and setting

The study is designed as a qualitative multiple-case study with embedded process analysis. Each student constitutes a case, while individual AI-mediated writing episodes constitute embedded units of analysis. A case-study approach is appropriate because AI use cannot be understood apart from the task, course expectations, students' language proficiency, prior technology experience, and evolving drafts. The design is process-oriented rather than perception-oriented: participants are followed while composing multiple texts, allowing patterns to be compared within and across cases.

The study was conducted in an undergraduate English Education program at an Indonesian university within a writing-intensive course in which students regularly produced academic texts. Data collection lasted six weeks. GenAI was not introduced solely for research purposes; all participants had prior experience with at least one AI-supported writing tool. The lecturer clarified permitted and prohibited forms of AI assistance before data collection, while students remained responsible for source verification, final wording, and compliance with course-level academic-integrity expectations.

3.2. Participants and sampling

Sixteen undergraduates were selected through purposive maximum-variation sampling. Eligibility required enrollment in the target writing course, prior experience using GenAI for English writing, willingness to share AI interaction logs and writing artefacts, and consent to screen recording and stimulated recall. Writing proficiency was classified using writing-course GPA together with an IELTS Writing equivalent: high proficiency (n = 5; GPA ≥ 3.70; IELTS-equivalent 6.5–7.0), intermediate proficiency (n = 7; GPA 3.00–3.69; IELTS-equivalent 5.5–6.0), and low-intermediate proficiency (n = 4; GPA < 3.00; IELTS-equivalent 4.5–5.0). All 16 participants used Grammarly Premium; 12 used ChatGPT-4o, nine used QuillBot, and five used Gemini. Tool counts overlap because most participants used more than one platform.

Table 1. Participant writing-proficiency profile and GenAI tool engagement

Dimension	Category / criterion	n	Percentage / note
Writing proficiency	High: GPA ≥ 3.70; IELTS Writing equivalent 6.5–7.0	5	31.3%
Writing proficiency	Intermediate: GPA 3.00–3.69; IELTS Writing equivalent 5.5–6.0	7	43.8%
Writing proficiency	Low-intermediate: GPA < 3.00; IELTS Writing equivalent 4.5–5.0	4	25.0%
Tool engagement	Grammarly Premium	16	100%
Tool engagement	ChatGPT-4o	12	75.0%
Tool engagement	QuillBot	9	56.3%
Tool engagement	Gemini	5	31.3%

3.3. Writing tasks and data corpus

Participants completed four connected English-writing tasks over six weeks: a problem-solution text, an argumentative essay plan, a 900–1,100-word argumentative essay, and a revision-and-reflection task. Across the four tasks, the dataset contained 64 task-specific final texts (16 students × 4 tasks). GenAI use was permitted within clearly stated boundaries for brainstorming, explanation, language feedback, counterarguments, and revision suggestions, while students remained responsible for verification, source use, and the final submission.

Data collection combined naturally occurring digital traces with reflective accounts. Screen recordings captured 48.5 hours of writing and AI interaction, averaging 3.03 hours per participant across the complete task set. Exported logs yielded 342 prompt-response exchanges that were segmented as decision episodes. The corpus also included 64 task-specific final texts, writing-history evidence visible in the composing sessions, semi-structured interviews, and stimulated-recall conversations anchored in selected screen-recorded episodes. This triangulation connected what students did on screen with what entered the final text and how they later explained decisions about ownership, verification, and integrity.

Table 2. Multimodal data corpus and analytical purposes

Data source	Observed volume	Analytical purpose
Task-specific final texts	64 texts (16 × 4 tasks)	Trace outcomes across repeated writing tasks
Screen recordings	48.5 hours; M = 3.03 hours/participant	Observe prompting, checking, transfer, and revision sequence
Prompt-response logs / decision episodes	342 exchanges	Identify AI-use patterns, learner evaluation, and agency

Data source	Observed volume	Analytical purpose
Semi-structured interviews	16 participants	Explore authorship, integrity, and perceived responsibility
Stimulated recall	Episode-based recall anchored in recordings	Explain acceptance, rejection, and transformation decisions

3.4. Analytical procedure

Analysis proceeds in five stages. First, each AI interaction is segmented into a decision episode consisting of a writing need, prompt, AI output, learner evaluation, and subsequent action. Second, episodes are coded by function: idea generation, planning/organization, language repair, elaboration, evidence support, revision feedback, or full-text generation. Third, uptake is coded as accepted verbatim, accepted with minor editing, transformed substantially, rejected, or used only as a trigger for new independent writing. Fourth, thematic analysis is applied to interviews and stimulated-recall data to identify recurring explanations related to trust, convenience, learning, ownership, disclosure, and responsibility. Fifth, cross-case matrices connect observable behaviors with participants stated beliefs and produce pattern profiles.

The five AI-mediated agency dimensions provide a sensitizing framework rather than a rigid deductive template. Purpose is coded when students can connect AI use to a specific writing need. Verification includes source checking, factual checking, comparison with other references, linguistic judgement, and task-criteria checking. Transformation includes synthesis, reorganization, personalization, and substantial rewriting. Disclosure includes awareness of when AI involvement should be acknowledged. Responsibility includes the ability to explain, defend, and revise final claims independently. New categories are retained when they do not fit these dimensions.

To strengthen trustworthiness, a second coder independently codes approximately 20% of decision episodes. Disagreements are discussed and the codebook refined before the full analysis. An audit trail preserves coding decisions and cross-case matrices. Negative-case analysis actively seeks episodes that contradict dominant patterns, such as high-frequency users who demonstrate strong agency or low-frequency users who delegate substantial work. Member reflection can be used at the interpretive stage, asking participants whether case summaries reasonably represent their practices without treating participant agreement as the sole criterion of validity.

3.5. Ethics

Because AI logs and screen recordings could contain personal information, the study applied data-minimization, redaction, and secure-storage procedures. Research participation was separated from normal course participation, and students were informed that declining or withdrawing from the research would not affect their academic standing. The study examined AI-mediated writing practices rather than conducting disciplinary surveillance. Identifiers were replaced with participant codes in the analytic corpus and reporting. The final submitted manuscript should insert the institution-specific ethics approval reference and any formal consent statement required by the target journal.

4. RESULTS

The corpus contained 342 AI-mediated decision episodes captured across four writing tasks per participant. The analysis did not treat each participant as belonging permanently to a single user type; instead, patterns were coded at the episode level because the same student could demonstrate high agency in one task and lower agency in another. Four recurring patterns accounted for all recorded episodes: selective scaffolding (132 episodes; 38.6%), iterative co-composition (100; 29.2%), surface repair (73; 21.3%), and delegated composition (37; 10.8%). Taken together, selective scaffolding and iterative co-composition represented 67.8% of all episodes and were characterized by comparatively stronger learner direction, evaluation, and transformation.

Table 3. Distribution of GenAI-use patterns and learner agency

GenAI-use pattern	Frequency (n)	Percentage	Agency level	Core behaviour
Selective scaffolding	132	38.6%	High	AI bounded to brainstorming, outline support, or context-specific lexical help
Iterative co-composition	100	29.2%	Moderate–high	Prompts modified recursively; claims challenged; AI and student text actively integrated
Surface repair	73	21.3%	Moderate	Grammar, punctuation, wording, and diction corrected without changing central argument
Delegated composition	37	10.8%	Low	Whole AI-generated paragraphs copied with little verification or substantive rewriting
Total	342	100%	—	Episode-level distribution

4.1. Pattern 1: Selective scaffolding — AI as a deliberately bounded resource

Selective scaffolding was the most frequent pattern, occurring in 132 episodes (38.6%). In these episodes, students deliberately restricted AI to bounded functions such as brainstorming, outlining, generating alternatives, or locating context-specific synonyms. The defining feature was not infrequent AI use but purposeful containment: the learner identified a problem, used AI to widen the available options, and then returned to independent composing.

Observed selective-scaffolding episodes typically began with a learner-defined writing need. Students requested several possibilities rather than a finished paragraph, compared suggestions with their existing argument, and either selected one element or used the output only as a cognitive trigger. This pattern showed high agency because AI remained subordinate to a student-defined rhetorical purpose.

> Stimulated-recall evidence from S04, an intermediate-proficiency participant, illustrates this boundary-setting logic. In Task 3, S04 prompted ChatGPT: "Give me 3 strong arguments why AI should be regulated in education with supporting academic sources." The screen recording showed that S04 rejected the second generated point because the proposed legal framing was considered ambiguous, then substantially rewrote the first argument to connect it with the preceding paragraph. S04 explained: "Saya ambil ide utamanya tentang privasi data, tapi kalimat ChatGPT terlalu kaku dan terkesan generik. Jadi saya rombak total strukturnya supaya sesuai dengan argumen yang sudah saya buat di paragraf sebelumnya. Kalau saya pakai langsung, tulisan ini bukan punya saya lagi."

This episode demonstrates how agency was preserved through rejection, restructuring, and ownership reasoning. The student did not equate AI assistance with authorship transfer; instead, ownership depended on retaining control over argument selection, rhetorical fit, and final formulation. Across selective-scaffolding episodes, this process-based view of integrity was more informative than a simple measure of whether AI had been used.

4.2. Pattern 2: Iterative co-composition — high interaction with sustained human direction

Iterative co-composition accounted for 100 episodes (29.2%). These episodes involved sustained interaction in which students repeatedly modified prompts, questioned the relevance or accuracy of AI responses, rejected problematic claims, and combined selected AI material with their own ideas. Agency was rated moderate to high because the interaction remained recursive and learner-directed rather than ending with the first generated answer.

The strongest marker of iterative co-composition was recursive evaluation. Students used follow-up prompting to narrow, challenge, or reframe earlier outputs and then integrated only selected components into the developing text. In contrast to delegated generation, the final textual direction emerged through a series of learner decisions that remained traceable across prompt histories and screen recordings.

Across these episodes, the analytic emphasis was placed on observable revision and evaluation behaviour rather than on self-reported frequency of AI use. High interaction volume did not automatically indicate dependency; frequent users could still maintain strong agency when they verified, transformed, and could account for major textual decisions.

This pattern complicates the assumption that extensive AI use is equivalent to dependency. The more defensible distinction is between interaction volume and decision ownership. Iterative co-composition remained compatible with authorship when students could trace, justify, and revise the decisions that shaped the final text.

4.3. Pattern 3: Surface repair — active editing but limited higher-order engagement

Surface repair appeared in 73 episodes (21.3%). AI use in this pattern was confined mainly to grammar, punctuation, sentence-level correction, paraphrasing, and vocabulary variation. Agency was coded as moderate because students generally retained control over the argument and content, but the interaction could become mechanically dependent when linguistic suggestions were accepted without adequate understanding of their effect on meaning, stance, or register.

The findings show that conceptual ownership and linguistic ownership are not identical. A student may remain the source of the underlying argument while ceding substantial control over expression. Consequently, responsible AI-supported writing requires students to evaluate not only whether a correction is grammatically acceptable but also whether it preserves intended meaning, degree of certainty, disciplinary register, and personal voice.

Interpretation of this pattern was grounded in coded screen-recording and prompt-response evidence. No additional verbatim quotation is reported here because only documented participant quotations were retained in the empirical manuscript.

4.4. Pattern 4: Delegated composition — convenience displacing judgement

Delegated composition represented 37 episodes (10.8%), the smallest but most academically consequential pattern. These episodes involved copying whole AI-generated paragraphs or substantial blocks of prose with little verification or substantive rewriting. Agency was coded as low because the student transferred not only linguistic formulation but also a substantial portion of conceptual organization to the system.

The key issue was not textual overlap alone but decision ownership. Cosmetic editing after full-paragraph generation did not necessarily restore authorship when students had not determined the argument structure, verified claims, or substantially transformed the generated reasoning. Delegated composition therefore represents a form of agency loss even when the final text contains minor student edits.

The delegated-composition episodes provided the clearest connection between over-reliance and academic-integrity risk. When students could not independently justify key claims or organizational decisions, the problem extended beyond disclosure to intellectual responsibility for the submitted text.

4.5. Cross-pattern distribution and learner agency

Table 4. Illustrative decision episode from participant S04

Stage	Evidence from S04 (Intermediate proficiency; Task 3)
Prompt input	"Give me 3 strong arguments why AI should be regulated in education with supporting academic sources."
Observed evaluation	Rejected generated point 2 because the proposed legal instrument was judged ambiguous.
Observed transformation	Rewrote generated argument 1 in the student's own language and restructured it to fit the preceding paragraph.
Stimulated-recall interpretation	Ownership was linked to retaining control over argument selection, organization, and final wording.

The distribution across 342 episodes shows why a binary AI-user/non-user distinction is analytically weak. A clear majority of episodes (232/342; 67.8%) fell into selective scaffolding or iterative co-composition, where students retained relatively strong control through bounded prompting, rejection, verification, and transformation. Surface repair accounted for 73 episodes (21.3%) and represented a more limited form of agency concentrated on language form. Delegated composition accounted for 37 episodes (10.8%) and marked the clearest shift from assistance to substitution. The central pedagogical issue is therefore the quality of decision-making surrounding AI uptake, not mere exposure to the technology.

4.6. A five-dimension AI-mediated writing agency framework

Table 5. Analytical indicators of AI-mediated writing agency

Dimension	High-agency indicators	Low-agency indicators
Purpose	Specific reason for AI use; learner defines problem	Open-ended delegation without a defined learning/writing need
Verification	Checks facts, sources, language, rubric fit; compares alternatives	Assumes fluency equals accuracy; no source or meaning checks
Transformation	Synthesizes, rewrites, personalizes, rejects, or recombines	Copies or lightly edits while retaining machine reasoning
Disclosure	Understands course rules; can document assistance when required	Conceals use or is uncertain about expectations
Responsibility	Can explain and defend claims, organization, and wording	Cannot account for major decisions in final text

5. DISCUSSION

The central argument emerging from the analysis is that AI use should be understood as a distribution of decision-making rather than as a yes/no behavior. This reframes the long-standing concern that technology may either "help" or "replace" the writer. In practice, both processes can occur within the same essay. A student may independently formulate a thesis, delegate a transition paragraph, critically verify a factual example, and automatically accept a vocabulary substitution. Agency is therefore episodic and uneven.

This finding extends earlier research on learner engagement with automated feedback. Fu and Liu (2022), Hoang (2022), Shang (2022), and Wang and Han (2022) all point, in different ways, to the importance of what learners do with machine-generated feedback. GenAI makes that engagement question more consequential because the technology can now generate content before and during drafting. The present framework suggests that future studies should code not only whether feedback is used but also who initiated the writing problem, how alternatives were evaluated, and who retained responsibility for the final rhetorical decision.

The four observed patterns also challenge frequency-based interpretations. Selective scaffolding and iterative co-composition can occur during frequent AI use while still showing substantial learner agency. The distinction lies in purpose, verification, and transformation. This matters for institutional policy. Rules based solely on frequency or on detecting AI-like language may misclassify responsible process use while failing to identify less visible delegation. A process-based approach to academic integrity would instead ask students to preserve traces, annotate AI contributions where required, or explain revision decisions.

Authorship in AI-mediated writing is best treated as accountable control rather than isolated word production. This does not mean wording is irrelevant. Language choices carry stance, certainty, identity, and disciplinary meaning. The surface-repair pattern demonstrates that students may preserve conceptual ownership while losing linguistic control. EFL pedagogy must therefore resist the assumption that grammar correction is pedagogically neutral. When a system changes modality, hedging, cohesion, or lexical register, students need enough language awareness to evaluate those changes.

The delegated-composition pattern reveals a second problem: editing is not equivalent to authorship. Students can heavily modify generated prose without constructing the underlying argument. Traditional plagiarism-oriented indicators often focus on text overlap, but GenAI can

produce original wording with weak human intellectual participation. This suggests that authentic assessment in the GenAI era should increasingly value process evidence: oral defense, version history, in-class planning, reflective commentary, source trails, and explicit rationale for major revisions.

The findings also support a more educational interpretation of academic integrity. Integrity should not be taught only as a list of prohibitions. Students need concrete practice in identifying acceptable assistance, verifying model outputs, preserving their own rhetorical purposes, and disclosing use appropriately. The five-dimension framework can become a classroom heuristic. Before accepting AI assistance, students can ask: What problem am I trying to solve? How will I verify this? What will I transform? What must I disclose? Can I defend the final result?

Finally, the study highlights the importance of AI literacy as evaluative literacy. Knowing how to prompt is not enough. Effective prompting can actually increase the amount of plausible but unverified content entering a draft. Students need epistemic strategies for checking claims and references, rhetorical strategies for judging fit with audience and genre, linguistic strategies for understanding revisions, and ethical strategies for determining when assistance becomes substitution.

5.1. Pedagogical Implications

Writing courses can operationalize responsible GenAI use through process-visible assignments. One approach is to require a short AI-use appendix in selected tasks that records the purpose of major prompts, identifies which suggestions were accepted or rejected, and explains one verification decision. This should not become bureaucratic surveillance; its purpose is metacognitive. Students learn to articulate their role in the writing process.

A second implication is to teach "critique before uptake." In classroom demonstrations, teachers can intentionally provide an AI-generated paragraph containing a mix of strong language, vague claims, false certainty, and weak evidence. Students evaluate it against a rubric, verify information, and decide what should be retained, transformed, or discarded. Such tasks make critical AI use a visible skill rather than an assumed competence.

Third, assessment should include moments where students must demonstrate ownership. Short conferences, oral explanations of argument structure, or annotated revision histories can provide richer evidence of authorship than AI detectors. Fourth, course policies should distinguish functions of AI use rather than merely naming tools. A policy can specify whether brainstorming, language feedback, paraphrasing, source discovery, and text generation are permitted, conditionally permitted, or prohibited for a particular assignment.

5.2. Limitations and Directions for Future Research

The study has several limitations. First, the single-program setting limits transferability, and replication across institutions, writing genres, and proficiency levels is needed to test the stability of the four episode-level patterns. Second, screen recording may influence composing behaviour, and students can use additional devices or offline resources that are not fully captured. Third, the pattern frequencies should not be interpreted as fixed learner identities because individual students may shift between high- and low-agency practices across tasks. Future research should therefore combine process traces, interviews, and artefact histories and examine how agency changes over time.

Future empirical work could quantify the framework after qualitative validation, develop an AI-mediated writing agency scale, or examine whether explicit instruction in verification and transformation changes students' writing behavior over time. Longitudinal research is especially needed to determine whether sustained GenAI use develops or erodes independent writing capacity.

6. CONLCUSION

The most productive research question about GenAI in EFL writing is no longer whether students perceive the technology positively. The more consequential question is how students

distribute cognitive and rhetorical work between themselves and the system. By tracing prompts, outputs, decisions, drafts, and explanations, this study design makes that distribution visible. The proposed framework defines responsible AI-mediated writing through purpose, verification, transformation, disclosure, and responsibility. These dimensions preserve room for legitimate technological support while keeping human judgement and accountable authorship at the center of writing pedagogy.

7. METHODOLOGICAL TRANSPARENCY AND REPLICABILITY NOTES

A publishable version of this study should preserve the process-oriented logic while tightening the connection between claims and observable evidence. The strongest contribution will come from showing not merely what participants say about AI but how specific writing episodes unfold. For each focal case, the final article should present a compact evidence chain: the writing problem identified by the student, the prompt submitted to the system, a short description of the AI response, the student's subsequent action, the change visible in the draft, and the participant's explanation during stimulated recall. This chain allows readers to judge the interpretation rather than accepting thematic labels on trust. It also makes transparent where the researcher infers agency and where that inference is directly supported by participant explanation.

Sampling should be purposeful rather than convenience-only. The actual dataset should include contrasting profiles: high and low writing achievement, frequent and occasional AI users, students who primarily use language-correction tools, and students who use generative systems for planning or content. Variation matters because a homogeneous sample can create a misleading single story of "the EFL student." During analysis, researchers should deliberately compare cases that reach similar final writing quality through different processes. For example, one student may produce a strong essay through selective AI consultation, while another may produce an equally polished essay through extensive generation followed by editing. The contrast helps separate product quality from authorship process.

The real study should also record course-level rules as data. Students' integrity decisions are shaped by what teachers permit, prohibit, or leave ambiguous. A short document analysis of assignment instructions, course syllabi, AI statements, and assessment rubrics can contextualize participants' decisions. If a lecturer says "AI may be used for language checking but not idea generation," the same prompt has a different ethical status than it would in a course that explicitly allows collaborative ideation. Academic integrity should therefore be interpreted relative to communicated expectations, not only the researcher's assumptions.

A useful reporting strategy is to distinguish descriptive codes from evaluative interpretations. "Requested complete paragraph," "copied 80% of response," and "checked cited source" are descriptive observations. "Low agency," "responsible use," or "delegated authorship" are higher-level interpretations. Keeping these levels separate reduces moral overreach. Researchers should also include negative cases. A participant who copies a long generated passage but rigorously verifies it may challenge one dimension of the framework; a participant who writes every sentence independently but uses AI to determine the entire argument may challenge another. Such cases refine the theory.

Finally, the article should avoid presenting AI-mediated agency as a fixed personality trait. The same student may show high agency in one task and weak agency in another depending on time pressure, topic knowledge, assessment stakes, or language difficulty. The unit of interpretation should therefore move between episode, task, and person. This multilevel view is especially valuable for pedagogy because it suggests that responsible practice can be taught. Students are not simply "good" or "bad" AI users; they develop repertoires of decisions that can become more critical, transparent, and purposeful through instruction.

For data saturation, the research team should not rely on a predetermined number alone. Instead, collection and analysis can proceed iteratively until additional cases no longer substantially extend the decision-pattern framework. Because screen data are highly detailed, sixteen cases may generate a very large corpus. A focused sampling strategy for stimulated recall

is therefore essential: select episodes representing high uptake, rejection, factual checking, transformation, and possible delegation rather than replaying every interaction. This keeps interviews cognitively manageable while preserving analytic depth.

The final manuscript should include at least two extended case vignettes alongside cross-case tables. Tables reveal patterns; vignettes reveal process. One vignette could show a high-agency student moving from a vague idea to a self-authored argument through selective prompting and verification. A contrasting vignette could show how convenience gradually shifts into delegation under deadline pressure. Together, these cases can make the central argument vivid: responsible GenAI use is not defined by tool presence but by who retains control over purposes, judgements, transformations, and accountability.

9.1. Reporting Transparency and Data Management

For the actual publication, data management should be planned before collection. Prompt logs and screen recordings can reveal names, account identifiers, pasted assignment text, or third-party information, so the research team should create a redaction protocol and a secure naming system. The manuscript should report exactly which forms of data were collected, how long they were retained, who had access, and how excerpts were selected. Where participants quote AI output, the authors should distinguish participant-generated language from machine-generated language in the analytic archive. A preregistered or publicly timestamped analysis plan is not mandatory for qualitative work, but documenting the planned research questions, sampling logic, core data sources, and initial sensitizing framework before full analysis can strengthen transparency while still allowing inductive development.

When publishing excerpts, researchers should avoid presenting long AI outputs that are unnecessary for the analytic point. Short, contextualized segments are sufficient to show the decision sequence. Participant quotations should be de-identified, and every verbatim quotation should be traceable to an audio transcript and participant code. The present manuscript reports only the documented S04 stimulated-recall quotation supplied with the dataset; other patterns are interpreted from coded behavioural evidence without reconstructing participant speech.

8. REFERENCES

Bandura, A. (2001). Social cognitive theory: An agentic perspective. Annual Review of Psychology, 52, 1–26.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Fu, H., & Liu, X. (2022). EFL learner engagement in automatic written evaluation. Frontiers in Psychology, 13, 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, 100055.

Hoang, G. T. L. (2022). Feedback precision and learners' responses: A study into ETS Criterion automated corrective feedback in EFL writing classrooms. The JALT CALL Journal, 18(3), 444–467.

Hyland, K. (2019). Second Language Writing (2nd ed.). Cambridge University Press.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic Inquiry. Sage.

Roe, J., & Perkins, M. (2022). What are automated paraphrasing tools and how do we address them? A review of a growing threat to academic integrity. International Journal for Educational Integrity, 18.

Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: A CALL for intelligent practice. Anglistik: International Journal of English Studies, 33(1), 165–184.

Shang, H.-F. (2022). Exploring online peer feedback and automated corrective feedback on EFL writing performance. Interactive Learning Environments, 30(1), 4–16.

UNESCO. (2023). Guidance for Generative AI in Education and Research. UNESCO.

Wang, S., & Xu, Y. (2022). Exploring the college EFL self-access writing mode based on automated feedback. International Journal of Contemporary Education, 5(1), 29–33.

Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research. Frontiers in Psychology, 13, 909802.

Woo, D. J., Wang, Y., & Susanto, H. (2023). Understanding EFL students' idea generation strategies for creative writing with natural language generation tools. [Bibliographic details to be verified against the source database before submission].

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods (6th ed.). Sage.

APPENDIX A. Proposed Semi-Structured Interview and Stimulated-Recall Protocol

Core interview domains: (1) Describe the stages of writing in which you usually use GenAI and why. (2) Tell me about a time you rejected an AI suggestion. What made you reject it? (3) How do you check whether generated information or references are accurate? (4) When does AI assistance still feel like "your own writing," and when does it not? (5) What kinds of AI use should be disclosed to a lecturer or reader? Why? (6) Has regular AI use changed what you can do without AI? (7) How do course rules influence your decisions?

Stimulated-recall prompts tied to screen clips: "What problem were you trying to solve here?" "What did you notice in the AI response?" "Why did you copy/change/reject this part?" "Did you verify anything before using it?" "Who made the key decision in this episode?" "Could you reproduce or defend this point without the AI?"

APPENDIX B. Decision-Episode Coding Template

Table 6. Decision-episode coding template

Field	Coding entry
Episode ID	Participant-task-sequence (e.g., P07-T3-E12)
Writing need	Problem recognized before prompting
Prompt purpose	Idea / plan / language / evidence / revision / generation
AI output type	Suggestion / explanation / rewritten text / full paragraph / source lead
Evaluation behavior	Check / compare / question / no visible evaluation
Uptake	Verbatim / minor edit / transformed / rejected / trigger-only
Verification	Factual / source / linguistic / rubric / none
Agency dimensions	Purpose / verification / transformation / disclosure / responsibility
Analytic memo	Interpretation and contradiction notes