IDEAS

Journal on Language Teaching and Learning, Linguistics and Literature

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

Volume 14, Number 2, December 2026
pp. 7340 - 7354

Issued by English study program of UIN Palopo

Prompt Literacy and Editing Behaviors in AI-Assisted Thesis Proposal Writing: A Multiple Case Study of Indonesian EFL Students

Rahma[1], Delita Sartika[2], Nely Arif[3]
[1,2,3]Magister Pendidikan Bahasa Inggris, Universitas Jambi, Kota Jambi, Jambi

Article Info

Received: 2026-09-01
Revised: 2026-09-01
Accepted: 2026-09-10

Keywords:
ChatGPT, prompt literacy, editing behaviors, EFL academic writing, human-AI interaction, thesis proposal writing

DOI:
10.24256/ideas.v14i2. 12511

Corresponding Author:
Rahma
rahmaaa1903@gmail.com
delitasartika@unja.ac.id
Magister Pendidikan Bahasa Inggris, Universitas Jambi, Kota Jambi, Jambi

Abstract

Research on AI-assisted writing in EFL contexts has been growing rapidly. However, most studies rely on self-reported perceptions, leaving a gap between what students say about their AI use and what they actually do. This qualitative multiple case study examined EFL students' perceptions of ChatGPT and their observable interaction patterns during the writing of a thesis proposal. The study involved three undergraduate EFL students from Universitas Islam Batang Hari. Data were gathered using semi-structured interviews with stimulated recall, full ChatGPT conversation logs shared through an export link, and a structured observation checklist. Interview data were analyzed thematically, while the documentary data 31 substantive prompts in total were analyzed using a six-category prompt classification taxonomy and a four-category editing behavior typology. Results for RQ1 revealed three perception themes: ChatGPT as cognitive scaffold, awareness of referential unreliability, and ambivalent writing confidence. Results for RQ2 indicated that generation prompts were the most common prompt type across all participants, with prompt sophistication associated with participants' experience level and language preference. Analysis of editing behavior indicated a predominant pattern of selective use and substantial modification, with direct adoption occurring sporadically under time pressure. Cross-analysis of interview and documentary data revealed meaningful, participant-specific discrepancies between reported and observed practice, particularly with respect to the degree of independent rewriting. These findings contribute to the growing literature on human-AI interaction in academic writing by showing that direct documentary evidence can complement, and at times challenge, self-reported accounts of AI-assisted writing, with implications for institutional AI pedagogy and policy.

1. Introduction

Academic writing, particularly thesis proposal writing, represents one of the most demanding intellectual tasks for students learning English as a Foreign Language (EFL). Unlike everyday communicative writing, thesis proposal writing requires the simultaneous management of disciplinary content, formal academic register, citation conventions, and coherent argumentation across extended texts. For EFL students working in a language that is not their mother tongue, these demands are compounded by linguistic challenges including limited vocabulary range, uncertainty about grammatical accuracy, and unfamiliarity with the conventions of English academic discourse (Bulqiyah et al., 2021).

The result is that many EFL students experience significant anxiety and avoidance behavior when confronted with major academic writing tasks, particularly at the proposal stage of thesis development. Against this backdrop, generative AI tools, most prominently ChatGPT, have emerged as resources that EFL students increasingly turn to for writing support. ChatGPT is capable of generating extended academic text, providing structural guidance, offering explanations of theoretical concepts, and producing language at a register appropriate for academic contexts. Its accessibility, speed, and continuous availability make it particularly attractive to students who lack immediate access to expert writing tutors or who feel unable to articulate ideas in formal English without support (Wale & Kassahun, 2024; Harunasari, 2023). Within a short period following its public release, ChatGPT has become a de facto writing assistant for substantial numbers of university students globally, and Indonesian EFL students are no exception to this trend.

Current research responding to this development has grown rapidly and can be organized around three interrelated strands. The first strand draws on the Technology Acceptance Model (TAM; Davis, 1989) as background to adoption-oriented perception research, rather than as the analytic framework adopted in the present study. Studies applying this framework consistently report that perceived ease of use positively influences perceived usefulness, which in turn shapes students' attitudes toward integrating ChatGPT into their writing workflow (Salam, 2024), with perceived usefulness in overcoming writer's block and producing academically appropriate text identified as the primary driver of adoption among Indonesian undergraduates (Sumakul et al., 2022). Broader surveys of higher education students corroborate generally positive attitudes toward generative AI for writing, research, and personalized learning support (Chan & Hu, 2023), although such general higher-education samples leave the specific ways language proficiency shapes EFL students' interaction patterns underexplored. Across this strand, however, perception is measured through surveys and interviews, without access to documentary evidence of what students actually do.

The second strand concerns prompt literacy: the ability to formulate inputs that elicit outputs relevant, accurate, and appropriately formatted for the user's purpose (Zamfirescu-Pereira et al., 2023). This literature has documented a widening taxonomy of request types: new-text generation, or "ghostwriting," in which the role of production is transferred wholesale to the AI (Jelson et al., 2025); revision requests, which preserve students' evaluative control over drafts they have already produced (Woo et al., 2025); clarification requests, which position students as active inquirers rather than passive recipients of content (Jelson et al., 2025); translation prompts, through which ChatGPT is used as a linguistic mediator between students' first language and English (Werdiningsih et al., 2024); structural prompts requesting outlines or chapter organization, associated with higher-achieving students who decompose tasks into multi-step sequences rather than single undifferentiated requests (Woo et al., 2025; Sawalha et al., 2024); and paraphrasing

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

prompts, a hybrid form of authorship in which students supply propositional content while delegating linguistic expression to the AI (Werdiningsih et al., 2024). Prompting patterns have further been shown to vary with students' AI literacy levels and to shape the character of the resulting learning experience (Kim et al., 2025).

The third strand examines what students do with AI-generated output once it is produced. Framing the writer as editor and curator rather than sole author (Ippolito et al., 2022), this literature has shown that editing behavior differs markedly by proficiency: higher-proficiency writers tend to use AI output as a structural scaffold, rewriting substantially to reflect their own voice and research context, or selectively extracting specific phrases and frameworks, while lower-proficiency writers more often adopt AI output directly with minimal modification (Chen et al., 2024). Critically, the same study finds that students' self-reports of their own editing behavior systematically overstate the extent of independent modification, a finding with direct methodological implications for how such behavior should be studied.

Taken together, these three strands demonstrate that ChatGPT use in EFL academic writing is well documented at the level of perception and, increasingly, at the level of behavioral typology. Yet a consistent limitation cuts across them: a substantial portion of existing research rests on self-reported perceptions, attitudes, and satisfaction surveys (Salam, 2024; Sumakul et al., 2022; Chan & Hu, 2023; Jelson et al., 2025) rather than on direct documentary evidence of interaction. What such studies cannot reveal is what students actually do when they sit with ChatGPT open and a blank proposal document in front of them: the specific prompts they write, the way they evaluate and modify AI-generated output, and the patterns that distinguish more sophisticated users from less sophisticated ones.

This gap is not merely methodological but substantive: self-reported accounts of AI use may diverge from actual practice for several distinct reasons, including social desirability bias, inaccurate recall of specific interactions, and a mismatch between students' generalized self-perception and their situation-specific behavior in a given writing session. Where researchers and educators base their understanding of student AI use on perception data alone, they risk overestimating the quality and criticality of that engagement, with direct consequences for evidence-based instructional design, institutional policy, and academic integrity frameworks.

The present study addresses this gap by examining, within a single design, both EFL students' perceptions of ChatGPT and their actual interaction patterns during thesis proposal writing. Distinct from prior perception-based work, it adopts a document-based approach that directly analyzes complete ChatGPT conversation logs alongside structured observation data and in-depth interview transcripts, allowing stated perception to be checked against documented behavior rather than treated as its proxy. Conceptually, the study treats prompt literacy, AI output, and editing behavior as interlocking stages of a single human-AI writing process moving from prompt literacy to AI output, to evaluation and editing behavior, to human agency and authorship, and ultimately to the final academic text rather than as parallel, unconnected phenomena.

The study is anchored in three research questions: (1) How do Indonesian EFL students perceive the usefulness of ChatGPT in their thesis proposal writing? (2) What interaction patterns, in terms of prompt types and editing behaviors, characterize EFL students' actual use of ChatGPT in thesis proposal writing? and (3) To what extent do students' reported AI-assisted writing practices converge with or diverge from their documented interaction patterns? Data were collected from three undergraduate EFL students at a university in Indonesia, each of whom shared their complete ChatGPT conversation logs for analysis alongside participation in stimulated recall interviews.

2. Method

This study employs a qualitative multiple case study design (Yin, 2018). A case study approach is appropriate because the study examines a bounded phenomenon, ChatGPT-assisted thesis proposal writing, across three individual cases, using multiple data sources to produce analytical depth that survey-based or large-sample approaches cannot achieve. The multiple case design enables cross-case comparison that reveals both patterns common to all participants and meaningful variation across them, producing what Yin (2018) terms replication logic: findings that hold across cases carry greater weight than single-case findings, while divergences illuminate the role of individual and contextual factors.

Three EFL undergraduate students at Universitas Islam Batang Hari (pseudonyms P-01, P-02, and P-04) participated in the study. Participants were recruited using purposive sampling (Creswell, 2014) based on four criteria: (a) they were currently engaged in writing a thesis proposal; (b) they were using ChatGPT as part of their writing process; (c) they were willing to share complete ChatGPT conversation logs; and (d) they were available for an in-depth interview. All three participants who met these criteria and agreed to take part were included in the study, so as to permit an in-depth, cross-case comparison. Table 1 presents the profile of each participant.

Participant	Profile
P-01	Thesis topic: Skimming/Scanning in Reading Comprehension. Eight months of ChatGPT experience. Prompts written in English, using structured and context-rich formulations (see Results for analysis of prompt sophistication).
P-02	Thesis topic: Students' Difficulties in English Listening. Five to six months of ChatGPT experience. Prompts written in English; detected a hallucinated reference in AI output (see Results for analysis of within-case progression in prompting).
P-04	Thesis topic: Effect of English Songs on Pronunciation. Two years of ChatGPT experience; prompts written in Indonesian with a casual register; previously used an alternative AI application (Cici); openly reported occasional direct copy-paste of AI output.

Table 1. Participant Profiles

Instrument

Three sources of data were used in the study, comprising one naturally occurring documentary record and two researcher-designed instruments. The first, and primary, source of behavioral evidence was the ChatGPT conversation log itself a documentary record rather than a researcher-designed instrument obtained as a complete, unedited export of each participant's interaction history, documenting every prompt submitted and every response received across all thesis-proposal writing sessions, covering the Introduction, Literature Review, and Methodology sections of each participant's proposal, or the section(s) each participant had completed at the time of data collection.

The second, researcher-designed instrument was a semi-structured interview guide incorporating stimulated recall procedures. The guide consisted of three parts: (a) opening questions eliciting participants' general perceptions of ChatGPT's usefulness and limitations in thesis proposal writing; (b) a stimulated recall segment in which specific prompts and corresponding outputs drawn from the participant's own conversation log were displayed to the participant, who was then asked to explain the reasoning behind the prompt, the evaluation of the output received, and any subsequent editing decision; and (c) closing questions addressing the affective dimensions of AI-assisted writing, such as confidence, anxiety, and perceived ownership of the resulting text. Interviews were conducted in a combination of Indonesian and English, reflecting participants' natural bilingual interaction style, and lasted 45–60 minutes each.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

The third, researcher-designed instrument was a structured observation checklist, developed by the researcher prior to data collection to enable systematic, line-by-line coding of each participant's conversation log. The checklist recorded four dimensions of interaction for every logged exchange: prompt behavior (the type of request made), output engagement (whether and how the participant responded to the AI's output within the conversation), post-output editing pattern (how the output was subsequently treated in the proposal document), and iteration pattern (whether and how a prompt was refined across successive turns). The coding categories used in this checklist correspond to the two analytical frameworks.

Data Collection

Data collection proceeded in four sequential steps to ensure that each source of evidence could be cross-referenced against the others. First, following informed consent, each participant was asked to export and share their complete ChatGPT conversation log via a shared export link, covering the full duration of their thesis proposal writing process up to the point of data collection. Second, the researcher conducted a preliminary review of each log using the structured observation checklist, coding every prompt-response pair and flagging exchanges that illustrated notable prompting or editing decisions for follow-up. Third, semi-structured interviews were scheduled and conducted individually with each participant, via WhatsApp voice or video call and, where feasible, in person; the stimulated-recall segment of each interview drew directly on the exchanges flagged during the preliminary log review, so that participants were asked to account for their own documented behavior rather than for AI use in the abstract. Fourth, the structured observation checklist for each participant was finalized after the interview, with codes revisited and, where necessary, revised in light of the participant's own explanation of a given prompt or editing decision. All interviews were audio-recorded with participants' consent and transcribed verbatim prior to analysis.

Data Analysis

Data analysis followed a two-track approach corresponding to the study's two research questions, followed by a cross-analysis stage integrating both tracks.

For RQ1 (perceptions), interview transcripts were subjected to thematic analysis following Braun and Clarke's (2006) six-phase procedure: familiarization with the data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the final analytic write-up. Codes were generated inductively from the transcripts and subsequently organized into themes relating to perceived usefulness, perceived limitations, and the affective dimensions of AI-assisted writing.

For RQ2 (interaction patterns), documentary data from the ChatGPT logs were analyzed using two purpose-built coding frameworks. The first is a six-category prompt classification taxonomy adapted from Jelson et al. (2025), Sawalha et al. (2024), and Werdiningsih et al. (2024), comprising generation prompts (requesting new text production), revision prompts (requesting improvement of existing text), explanation prompts (requesting conceptual clarification), translation prompts (requesting language conversion), structural prompts (requesting organizational frameworks), and paraphrasing prompts (requesting rewording). Every prompt submitted by every participant was assigned to exactly one category; where a single prompt performed more than one function simultaneously (for example, requesting revision and translation in the same turn), it was coded according to its dominant communicative function as judged from the immediate conversational context, with the secondary function noted in the coding memo for that exchange.

The second is a four-category editing behavior typology drawn from Ippolito et al. (2022) and Chen et al. (2024): direct adoption (AI output used with minimal or no modification), substantial modification (AI output used as a scaffold but significantly rewritten), selective use (specific elements extracted and integrated into independently written text), and rejection (AI output not incorporated into the proposal). Coding against both frameworks was performed by the first author at the level of each individual prompt-response pair in the log, and subsequently verified through cross-referencing with the structured observation field notes completed for the same exchange. Throughout the Results, editing-behavior categories are reported together with their evidential basis whether directly observed in the log, self-reported in the stimulated-recall interview, or inferred from the interaction pattern-rather than being treated as uniformly observed facts; because final proposal drafts were not analyzed in this study, none of the editing categories independently verify how AI-generated text was ultimately incorporated into participants' submitted documents.

Cross-analysis was conducted as the final analytical step, in which the themes generated from the interview data (RQ1) were systematically compared against the coded patterns identified in the documentary log data (RQ2) for the same participant. This triangulation procedure was designed specifically to identify convergences and divergences between participants' stated perceptions and their observable behavior, the perception-behavior gap that constitutes the study's central analytical contribution

3. Results

RQ1: Perceptions of ChatGPT in Thesis Proposal Writing

Thematic analysis of interview data yielded three themes characterizing participants' perceptions of ChatGPT in their thesis proposal writing: (1) ChatGPT as cognitive scaffold, (2) awareness of AI limitations and reference unreliability, and (3) ambivalent writing confidence.

Theme 1: ChatGPT as Cognitive Scaffold

All three participants perceived ChatGPT primarily as a tool that reduces the cognitive load of academic writing initiation. The most consistently cited benefit across all participants was ChatGPT's capacity to provide a starting point a structural scaffold or initial draft at moments when participants experienced writer's block or did not know how to begin a section. P-01 described ChatGPT as an assistant that was particularly helpful when facing complex revision requests from her supervisor:

Ketika saya diminta oleh pembimbing untuk menambahkan konteks Indonesia dalam latar belakang saya, saya tidak tahu bagaimana cara mengintegrasikannya dengan paragraf yang sudah ada. Saya meminta ChatGPT untuk merevisi dua paragraf pertama saya dan hasilnya sangat memuaskan. (P-01, Q4)

P-02 emphasized ChatGPT's value for bridging the gap between conceptual understanding and written expression in English:

> *"I understand these theories from reading, but I could not explain them clearly in English. I gave ChatGPT my prompt with the three theories I wanted to include, and it wrote a well-organized literature review with subheadings and citations" (P-02, Q4)*

P-04 similarly highlighted ChatGPT's capacity to generate both ideas and structure simultaneously, describing it as helpful for mencari ide, topik, parafrase (finding ideas, topics, paraphrasing) (P-04, Q2). The perception of ChatGPT as a scaffold something that supports the student's own thinking rather than replacing it was the dominant self

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

conception across all three participants, though this self perception requires examination in light of the behavioral data discussed.

Theme 2: Awareness of AI Limitations and Reference Unreliability

A second consistent theme was participants' awareness that ChatGPT outputs are not uniformly reliable, with reference hallucination emerging as the most salient specific concern. P-01 described discovering that references provided by ChatGPT could not be verified: "Satu kali ChatGPT memberikan referensi akademik yang terlihat valid tetapi ketika saya cek di Google Scholar, referensi tersebut tidak ada" (P-01, Q6). This experience led P01 to adopt a systematic reference verification practice. P-02 reported a more specific incident: ChatGPT cited "Vandergrift (2003)" when the correct publication date is 2004 a small but consequential error that P-02 only caught through careful cross checking against Google Scholar. P-04 noted more generally that ChatGPT outputs were sometimes "ngak nyambung" (irrelevant or off topic) (P-04, Q6), requiring revision or reprompting. Across all three participants, these experiences of unreliability generated active verification behaviors rather than abandonment of the tool, suggesting that critical AI engagement can be learned through corrective experience rather than formal instruction.

Theme 3: Ambivalent Writing Confidence

The third theme reflects the complex affective relationship participants developed with ChatGPT over time. All three participants reported increased confidence in their capacity to produce academic text in English an outcome consistent with TAM's perceived usefulness construct. P-02 to summarize this briefly: "Before ChatGPT, I was very afraid of writing the literature review because I thought my English was not good enough. Now I feel I can do it" (P-02, Q7). However, this increased confidence coexisted with a concern, voiced by P-01 most explicitly, about the authenticity of the writing: "Kadang saya khawatir apakah tulisan saya benar-benar mencerminkan pemahaman saya sendiri atau hanya tulisan yang dihasilkan AI" (P-01, Q7). P-04 expressed a related ambivalence: "penggunaan ChatGPT membuat saya lebih percaya diri...kadang kadang juga membuat saya kurang yakin kalo cuma mengandal chatgpt saja" (P-04, Q7). This ambivalence suggests that participants maintained metacognitive awareness of ChatGPT's role in their writing process, even as they became increasingly reliant on ChatGPT for particular writing tasks.

RQ2: Interaction Patterns, Prompts and Editing Behaviors

Prompt Patterns

Table 2 presents the distribution of prompt types across the three participants, based on systematic analysis of their complete ChatGPT conversation logs. A total of 31 substantive prompts were identified and classified across all three participants, excluding one social greeting submitted by P-04 that carried no academic task content.

Table 2. Prompt Type Distribution Across Participants

Prompt Type	P-01	P-02	P-04	Total	%
Generation	5	5	5	15	48.4%
Revision	3	2	3	8	25.8%
Structural	1	1	2	4	12.9%
Explanation	1	1	1	3	9.7%
Translation	0	0	1	1	3.2%
Paraphrasing	0	0	0	0	0%
Total Prompts	10	9	12	31	100%

Generation prompts were the dominant prompt type across all participants, accounting for 48.4% of all prompts submitted (n = 15). This finding indicates that students' primary mode of AI engagement was requesting ChatGPT to produce new text from scratch,

rather than refining, checking, or elaborating text the student had already written. When nearly half of all interaction is generation-based, ChatGPT is called on more often to produce new text than to refine text students have already written; whether this constitutes scaffolding or substitution depends on how that generated text was subsequently evaluated and edited, a question taken up together with the editing-behavior findings below and in the Discussion.

Revision prompts formed the second largest category at 25.8% (n = 8), indicating that students did engage in some degree of iterative refinement submitting prior outputs back to ChatGPT with requests to expand, shorten, or improve specific sections. Structural prompts accounted for 12.9% (n = 4), reflecting moments when students requested organizational scaffolding such as outlines, additional sections, or framework paragraphs. Explanation prompts (9.7%, n = 3) captured instances where students asked ChatGPT to clarify a concept before proceeding to write about it. Translation prompts were the least frequent category (3.2%, n = 1), appearing exclusively in P-04's log. Paraphrasing prompts were not observed in the dataset. This absence is consistent with the overall pattern of generation dominant interaction: participants engaged ChatGPT primarily to produce text, not to refine text they had independently written, leaving paraphrasing which presupposes prior independent authorship outside the scope of their logged interactions.

Across the three participants, prompt sophistication varied considerably, and this variation is associated with-though not straightforwardly explained by-their ChatGPT experience levels and their chosen language of interaction; as shown below, P-04's case in particular indicates that experience duration alone does not reliably predict prompt sophistication. P-01 consistently produced the most elaborate and context-rich prompts, providing her identity as an EFL student, her thesis topic, the specific section required, the desired paragraph count, and the Indonesian EFL context she wanted incorporated. Her opening prompt in Session 1 illustrates this pattern:

Hi ChatGPT, I am an EFL student and I am writing my thesis proposal. My topic is about the use of skimming and scanning strategies in reading comprehension. Can you help me write the background of study for my introduction? Please make it around 3–4 paragraphs and suitable for an undergraduate thesis proposal. (P-01, Session 1, Prompt 1)

This prompt exemplifies what Zamfirescu-Pereira et al. (2023) describe as directive prompting: explicit task specification that constrains the AI's output space and reduces the likelihood of generic or contextually irrelevant responses. Crucially, P-01 also submitted the highest proportion of revision prompts relative to her total (3 out of 10), consistently returning to prior outputs to refine specific aspects rather than simply accepting them and moving on.

P-02 demonstrated a clear within-case progression in prompting sophistication across the session, representing the most analytically compelling case of such progression in the dataset. Her opening prompt was markedly brief and context-free "help me write introduction for my thesis about students difficulty in listening english" (13 words, no specification of student level, institutional context, theoretical orientation, or desired academic register). However, her second prompt already showed a substantial increase in specificity, identifying the target population (Indonesian university EFL students), three specific difficulty factors, the desired register (formal academic language), and preferred references (Vandergrift and Renandya). By Session 3, P-02 submitted what is arguably the most sophisticated prompt in the entire dataset a metacognitive evaluation request asking ChatGPT to assess whether her methodology was "complete and strong enough" and identify what was missing. This type of prompt, in which the student positions ChatGPT as

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

a critical reviewer rather than a text generator, is associated with the highest level of human-AI collaborative interaction (Kim et al., 2024).

P-04's prompt patterns are qualitatively distinct from those of P-01 and P-02 in two important respects. First, P-04 interacted with ChatGPT primarily in Bahasa Indonesia rather than English, using a casual, colloquial register that included informal address forms ("bro") alongside substantive academic task instructions. Second, P-04's session covered the broadest scope from initial title generation through background writing, paragraph-level elaboration, reference list production, research instrument development, and final English-language synthesis making her log the most structurally ambitious of the three. The final prompt, requesting consolidation of all prior outputs into a single English-language chapter ("tolong gabungkan semua nya dari atas sampai bawah menggunakan bahasa Inggris dengan grammar yang benar"), represents a delegation of both synthesis and language production to ChatGPT that is not paralleled in the other two cases.

The language choice in prompting carries particular analytical weight. P-01 and P-02 both prompted in English, meaning that the linguistic process of articulating an academic task in the target language was itself part of their writing activity. P-04, by prompting in Indonesian, effectively delegated the entirety of English academic text production to ChatGPT, retaining only the role of content director. This distinction resonates with Chen Z's (2024) finding that the relationship between a writer and AI output is partly determined by the degree to which the writer engages linguistically with the task, rather than merely directing it from outside the language. P-04 delegated a comparatively larger proportion of English-language text production to ChatGPT than the other two participants, which has implications for language development that are discussed in Section 5.

Editing Behaviors

Table 3 presents the distribution of editing behaviors across participants, based on cross-analysis of ChatGPT logs, observation checklists, and interview accounts; because these sources differ in what they can establish, the narrative below identifies, for each participant, which categories are directly evidenced in the log or observation data and which rest primarily on self-report (for example, P-04's occasional direct adoption is self-reported rather than independently observed).

Table 3. Editing Behavior Distribution Across Participants

Editing Behavior	P-01	P-02	P-04
Direct Adoption	Not observed	Not observed in log	Occasionally (self-reported)
Substantial Modification	Dominant	Dominant	Present
Selective Use	Present	Present	Dominant
Rejection	Not observed	Not observed	Not observed

P-01's editing behavior was characterized by a consistent pattern of substantial modification combined with selective use. Interview data confirm that P-01 read each ChatGPT output in its entirety before taking action, evaluated the relevance and validity of references, rewrote content in her own words while retaining academic phrasing, and submitted follow-up prompts when initial outputs did not meet her needs. Her description of this process is representative: "Biasanya saya membaca dulu seluruh output, kemudian saya pilih bagian-bagian yang menurut saya paling relevan dan tepat. Setelah itu saya tulis ulang dengan kata-kata saya sendiri, tapi tetap mengacu pada struktur dan ide yang diberikan ChatGPT" (P-01, Q10). Observation data corroborate this account: direct copy-paste behavior was not evidenced in the available log and interview data, and P-01's follow-up prompts consistently demonstrated comprehension of and engagement with prior outputs.

P-02's editing behavior showed a developmental trajectory across the session. Early interactions with ChatGPT, particularly the first session, showed a pattern of selective extraction taking well-constructed phrases and integrating them into surrounding text with relatively limited independent rewriting. As the session progressed, particularly following the hallucinated reference incident (Vandergrift 2003 vs. 2004), P-02's engagement became more critically evaluative. The student began explicitly contextualizing ChatGPT's generalized output to the specific setting of Universitas Islam Batang Hari students a form of adaptive editing that required both reading comprehension and contextual awareness. Interview data confirm this pattern: "I think of ChatGPT's output as a first draft that I then improve and personalize" (P-02, Q10), and "Basically I act as the final editor ChatGPT drafts, I decide" (P-02, Q13).

P-04's editing behavior presents the most analytically complex profile. Interview data at Q10 include an explicit acknowledgment of occasional direct adoption: "Saya kadang tidak langsung menyalin jawabannya. Saya membaca terlebih dahulu, lalu menulis ulang menggunakan kata-kata saya sendiri...dan kadang juga kalo mendesak dan merasa sudah cocok saya tinggal copy saja" (Sometimes I don't copy the answer directly. I read first, then rewrite in my own words...but sometimes when I'm pressed for time and it seems right, I just copy). The log confirms that P-04 submitted a final synthesis prompt requesting that all prior outputs be consolidated into a complete English-language background chapter, which could plausibly be adopted with minimal additional editing. Observation data note that P-04 did engage in targeted extraction of specific paragraphs for further elaboration, indicating that some critical reading occurred; however, the overall pattern is more consistent with selective use with occasional direct adoption than with the systematic substantial modification characteristic of P-01.

The Perception Behavior Gap

Cross-analysis of interview data against documentary evidence from the logs reveals meaningful gaps between participants' stated practices and their observable behaviors. This perception-behavior gap manifests differently across participants but is present in all three cases.

For P-01, the gap is relatively minor but notable. P-01 consistently described her practice as one of substantial rewriting and critical verification, and the log data broadly support this account. However, P-01's interview at Q15 reveals an important qualification: "kalo malas kadang tu saya gak banyak modifikasi lagi bahasanya yang dari output Chatgpt itu". This acknowledgment of context-dependent variation between her typical careful practice and more expedient behavior under low motivation is absent from the more idealized account she offered in response to direct questions about her editing process.

For P-02, the gap is most clearly visible in the relationship between self-reported independence and actual generation patterns. P-02 stated that her Research Questions were written entirely independently "those must come from my own analysis of the problem...I want those to be completely my own thinking" (P-02, Q11) and this specific claim is consistent with the log, which shows no generation prompt specifically for research questions. However, P-02's description of her overall editing process as consistently involving substantial independent rewriting is complicated by the log evidence showing that substantial sections of the Background of Study and Literature Review were generated by ChatGPT and adopted with primarily contextual rather than structural modification.

For P-04, the gap between stated and observed practice is most pronounced. While acknowledging occasional direct copy-paste (Q10), P-04 also described her overall practice in terms that suggest more systematic critical engagement than the log fully corroborates: "Saya memilih bagian yang relevan dengan topik penelitian saya dan memeriksa apakah

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

informasinya sesuai dengan sumber akademik" (P-04, Q13). The log, however, shows that verification behavior while present in principle was less systematic than this description implies, and that the final consolidation prompt requesting a complete English synthesis of all prior outputs represents a form of delegation that is inconsistent with the active editorial role P-04 describes. This gap is the most analytically significant in the dataset, precisely because it illustrates the distortion that perception-only research methods introduce: P-04's self-report aligns with what a responsible student would ideally do, but the log documents a more expedient and less critically engaged practice.

4. Discussion

The findings of this study give rise to five interconnected discussion points that situate the results within the existing literature and draw out their implications for pedagogy and institutional practice.

First, the prompt patterns documented in this study provide empirical support for Zamfirescu-Pereira et al.'s (2023) account of prompt literacy as a developmentally acquired competence. The contrast between P-01's consistently elaborate, context-specific prompts and P-02's opening single-sentence prompt with P-02 subsequently developing more sophisticated prompting across sessions illustrates the trajectory from novice to developing prompt literacy that Zamfirescu-Pereira et al. describe. Critically, the data suggest that this development is not simply a function of experience duration: P-04, despite the longest ChatGPT experience in the sample (approximately two years), showed prompt patterns that were directive but not always strategically sophisticated, and that delegated substantial portions of the English language writing process to ChatGPT.

This suggests that prompt literacy development may be domain-specific P-04's prior experience was primarily with content generation tasks, and the particular demands of English academic writing had not been fully integrated into her prompting strategy. With only three cases, this pattern is best read as a proposition emerging from the data rather than a generalized conclusion. For pedagogy, this implies that prompt literacy instruction should be embedded within academic writing courses rather than treated as a general-purpose digital skill, as the relevant competence is not just knowing how to use AI, but knowing how to use AI specifically for academic writing purposes.

Second, the editing behavior spectrum documented in this study extends Chen Z., et al's (2024) proficiency-related findings by pointing to an additional factor that merits direct examination in future research. While Chen Z., et al find that lower-proficiency writers show higher rates of direct adoption, the cases here suggest that, alongside proficiency, the degree to which the student positions themselves as a critical agent in the writing process may also matter. P-01 and P-02, who both adopted a systematic reading-evaluation-modification workflow, showed minimal direct adoption regardless of their proficiency level. P-04, who had the most ChatGPT experience but the least systematic critical engagement, showed the highest rate of direct adoption not evidently because of lower proficiency, but because of a more expedient relationship to AI output. This finding resonates with Ippolito et al.'s (2022) argument that the pedagogically significant distinction is between students who position themselves as editors and curators of AI output and those who position themselves as recipients. The implication for writing instruction is that students need not only to be taught to use AI, but to be explicitly positioned as critical agents who bear authorial responsibility for every word in their final text.

Third, and most centrally, the perception-behavior gaps documented in Section 4.2.3 support and extend the critique that perception-only research methods risk overestimating the quality of student AI engagement. In all three cases, participants described their AI use

in terms that emphasized critical reading, independent rewriting, and systematic verification practices consistent with responsible, academically appropriate AI use. The documentary evidence from the logs reveals that, while these practices were genuinely present to varying degrees, they were also more inconsistent, context-dependent, and in P-04's case more limited than the interview accounts suggested.

This finding is consistent with what Chen Z., et al (2024) describe as systematic self-report bias in student accounts of AI use, and it has direct methodological implications for the field: with only three strategically selected cases, the appropriate claim is not that perception-only studies necessarily produce a false picture, but that the cases demonstrate self-reported accounts may not fully capture the variability and context-dependence of students' AI-assisted writing practices. The contribution of the present study lies precisely in its demonstration that document-based analysis of actual interaction logs provides a corrective lens not because students are being deliberately dishonest, but because self-perception and actual behavior regularly diverge, particularly in domains where social desirability pressures are active.

It should be acknowledged, however, that the ChatGPT logs analyzed here are themselves an incomplete record of participants' "actual writing practices": they document AI interaction but not what students subsequently typed or deleted in their Word or Google Docs drafts, what supervisor feedback shaped later revision, how much AI-generated text ultimately entered the submitted proposal, or how students revised offline without the AI. The logs therefore complement, rather than replace, self-reported accounts, and the perception-behavior gap reported here should be read as a gap between two partial, differently limited sources of evidence rather than between a flawed self-report and a complete behavioral record.

Fourth, the findings have implications for institutional policy on AI use in academic writing. The evidence from this study suggests that neither blanket prohibition nor unrestricted permissiveness represents an adequate institutional response to ChatGPT use in thesis writing. Prohibition is likely to drive AI use underground rather than eliminating it, while unrestricted permissiveness without pedagogical scaffolding produces the kind of expedient, minimally critical engagement documented in P-04's case. What the evidence supports, instead, is a framework that develops prompt literacy as a component of academic writing instruction and builds reflective awareness of the distinction between AI scaffolding and AI substitution.

Given that complete ChatGPT logs may contain unrelated personal exchanges, sensitive research ideas, or supervisor comments, and given the privacy, consent, and feasibility concerns that a mandatory submission policy would raise, this study's findings support voluntary or task-specific AI-use documentation such as short reflective AI-use statements, process portfolios, or selective disclosure of the specific exchanges relevant to a given section rather than mandatory submission of complete conversation histories. The within-case progression visible in P-02's data from vague early prompting to more strategic, critically engaged interaction suggests that this kind of growth is achievable within a single writing project, and that pedagogical support at the right moments could accelerate it.

Fifth, the findings raise a broader question of academic authorship and agency that cuts across the prompt and editing-behavior findings reported separately above. Read together, the cases suggest a continuum running from AI as scaffold, through AI as collaborator and AI as text generator, to AI substitution, along which individual prompt-response exchanges and indeed individual participants can be located. A sophisticated prompt does not by itself guarantee responsible authorship: a well-specified generation

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

prompt followed by extensive critical rewriting, as in P-01's case, can still represent strong writer agency, while a simple prompt adopted with minimal modification moves an exchange further toward substitution regardless of how the prompt was phrased.

Considered together with the editing-behavior findings, the cases point toward a working model in which prompt strategy leads to AI-generated output, which is then subject to evaluation and verification, followed by selection, revision, or rejection, which in turn determines the degree of writer agency and, ultimately, ownership of the final text. Within this model, verification behavior appears to play a particularly important role: P-01 and P-02's encounters with unreliable or hallucinated references appear to have triggered more careful, critical engagement with subsequent AI output, suggesting a possible mechanism running from an AI error encounter, to awareness of unreliability, to verification, to strengthened critical AI literacy. This mechanism, if supported by future research designed to test it directly, would be more theoretically informative than simply noting that students are aware ChatGPT can produce incorrect information.

5. Conclusion

This study examined how three Indonesian EFL undergraduate students perceived and actually used ChatGPT in the process of writing their thesis proposals, using a qualitative multiple case study design that integrated interview data with direct analysis of complete ChatGPT conversation logs and structured observation. The findings reveal three perception themes ChatGPT as cognitive scaffold, awareness of reference unreliability, and ambivalent writing confidence alongside documented interaction patterns showing that generation prompts dominated student-AI interaction, prompt sophistication varied significantly across participants, editing behaviors existed on a spectrum from selective use to occasional direct adoption, and meaningful gaps existed between participants' stated practices and their observable behaviors.

The study's primary contribution is methodological and substantive by moving beyond perception based inquiry to direct analysis of AI interaction logs, it provides evidence that self report data systematically overestimates the criticality and independence of student engagement with AI in academic writing. This finding challenges the evidential basis of much existing research in this area and calls for methodological diversification in future studies. Practically, the findings suggest that prompt literacy development should be integrated into academic writing instruction, and that institutional AI policies should be built on evidence of what students actually do rather than what they report doing.

Several limitations should be acknowledged. The study involves three participants from a single institution, which limits the generalizability of findings. The decision to analyze logs rather than final draft texts means that the ultimate integration of ChatGPT outputs into submitted proposals could not be directly verified. The interview based stimulated recall data, while valuable, remain susceptible to retrospective rationalization. Future research should pursue longitudinal designs that track prompt literacy development over extended periods, comparative studies across proficiency levels and institutional contexts, and automated approaches to prompt classification that could enable larger-scale analysis of student AI interaction logs.

6. References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa https://doi.org/10.1191/1478088706qp063oa

Bulqiyah, S., Mahbub, M. A., & Nugraheni, D. A. (2021). Investigating writing difficulties in essay writing: Tertiary students' perspectives. English Language Teaching

Educational Journal, 4(1), 61–73. https://doi.org/10.12928/eltej.v4i1.237 https://doi.org/10.12928/eltej.v4i1.237

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20(43). https://doi.org/10.1186/s41239-023-00411-8https://doi.org/10.1186/s41239-023-00411-8

Chen, Z., Zhu, X., Lu, Q., & Wei, W. (2024). L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. Computer Assisted Language Learning. 1-20. https://doi.org/10.1080/09588221.2024.2422478https://doi.org/10.1080/09588221.2024.2422478

Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008 https://doi.org/10.2307/249008

Harunasari, S. Y. (2023). Examining the effectiveness of AI-integrated approach in EFL writing: A case of ChatGPT. International Journal of Progressive Sciences and Technologies (IJPSAT), 39(2), 357–368. https://doi.org/10.52155/ijpsat.v39.2.5516https://doi.org/10.52155/ijpsat.v39.2.5516

Ippolito, D., Yuan, A., Coenen, A., & Burnam, S. (2022). Creative writing with an AI-powered writing assistant: Perspectives from professional writers. arXiv. https://doi.org/10.48550/arXiv.2211.05030 https://doi.org/10.48550/arXiv.2211.05030

Jelson, A., Manesh, D., Lee, S., Jang, A., Dunlap, D., Maddox, T., Kim, Y.-H., & Lee, S. W. (2026). NIRVANA: A comprehensive dataset for reproducing how students use generative AI for essay writing. arXiv. https://arxiv.org/abs/2604.07344.https://arxiv.org/abs/2604.07344

Jelson, A., Manesh, D., Jang, A., Dunlap, D., Kim, Y.-H., & Lee, S. W. (2025). An empirical study to understand how students use ChatGPT for writing essays. arXiv. https://arxiv.org/abs/2501.10551https://arxiv.org/abs/2501.10551

Kim, J., Yu, S., Lee, S.-S., & Detrick, R. (2025). Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. Journal of Research on Technology in Education, 57(1), 1–18. https://doi.org/10.1080/15391523.2025.2456043https://doi.org/10.1080/15391523.2025.2456043

Salam, U. (2024). The integration of ChatGPT in English for foreign language course: Elevating AI writing assistant acceptance. Computers in the Schools, 42(2), 145–165. https://doi.org/10.1080/07380569.2024.2446239https://doi.org/10.1080/07380569.2024.2446239

Sawalha, G., Taj, I., & Shoufan, A. (2024). Analyzing student prompts and their effect on ChatGPT's performance. Cogent Education, 11(1), Article 2397200. https://doi.org/10.1080/2331186X.2024.2397200 https://doi.org/10.1080/2331186X.2024.2397200

Sumakul, D. T. Y. G., Hamied, F. A., & Sukyadi, D. (2022). Students' perceptions of the use of AI in a writing class. Advances in Social Science, Education and Humanities Research, 624, 52–57.

ISSN 2338-4778 (Print)
ISSN 2548-4192 (Online)

https://doi.org/10.2991/assehr.k.220201.009https://doi.org/10.2991/assehr.k.220201.009

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. Management Science, 46(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926https://doi.org/10.1287/mnsc.46.2.186.11926

Wale, B. D., & Kassahun, Y. F. (2024). The transformative power of AI writing technologies: Enhancing EFL writing instruction through the integrative use of Writerly and Google Docs. Human Behavior and Emerging Technologies, 2024, Article 9221377. https://doi.org/10.1155/2024/9221377.https://doi.org/10.1155/2024/9221377

Werdiningsih, I., Marzuki, Indrawati, I., Rusdin, D., Ivone, F. M., Basthomi, Y., & Zulfahreza. (2024). Revolutionizing EFL writing: Unveiling the strategic use of ChatGPT by Indonesian master's students. Cogent Education, 11(1), Article 2399431. https://doi.org/10.1080/2331186X.2024.2399431https://doi.org/10.1080/2331186X.2024.2399431

Woo, D. J., Guo, K., & Susanto, H. (2025). EFL secondary students' use of ChatGPT for writing task completion pathways. The Journal of Educational Research, 118(6), 596–609. https://doi.org/10.1080/00220671.2025.2510382https://doi.org/10.1080/00220671.2025.2510382

Yin, R. K. (2018). Case study research and design methods (6th ed.). SAGE.

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Article 437, 1–21. https://doi.org/10.1145/3544548.3581388https://doi.org/10.1145/3544548.3581388