

Student Engagement with genAI Feedback in L2 Writing: Evidence from Chinese Students' Engagement with Kimi Feedback

Xiaohua Yang*, Xiumei Liu

Department of Foreign Languages, China University of Mining and Technology-Beijing, Beijing, China
Email: *meiller@163.com

How to cite this paper: Yang, X. H., & Liu, X. M. (2026). Student Engagement with genAI Feedback in L2 Writing: Evidence from Chinese Students' Engagement with Kimi Feedback. *Creative Education*, 17, 1739-1764.
<https://doi.org/10.4236/ce.2026.179099>

Received: July 20, 2026

Accepted: September 13, 2026

Published: September 16, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Despite substantial literature on the application of generative AI (genAI) feedback in L2 students' writing process, little has explored how students engage with such humanoid feedback. Informed by the multi-dimensional conceptual framework of student engagement, this study explores Chinese EFL students' affective, cognitive, and behavioral involvement in Kimi feedback. Multiple sources of data including archived Kimi records, multiple drafts, reflection papers, questionnaire, and interview were analyzed quantitatively and qualitatively. The results revealed the students exhibited overall positive engagement with Kimi's multi-leveled feedback. Affectively, they exhibited interest in Kimi feedback, happy to receive it, willing to negotiate with Kimi, and acknowledged the value of Kimi feedback in improving writing quality. Cognitively, they understood most Kimi feedback, deployed various cognitive and metacognitive operations to process the feedback, decide, and evaluate its incorporation. Behavioral engagement was found in high adoption of accurate Kimi feedback as well as dismissal of most inaccurate feedback. Meanwhile, behavioral engagement sometimes might emerge without sufficient cognitive engagement. The research provides evidence for the multi-faceted nature of student engagement with genAI feedback in L2 writing activity. It also highlights a multi-faceted lens for teachers to scaffold students for more effective utilization of such affordance.

Keywords

Student Engagement, Generative AI Feedback, EFL Writing

1. Introduction

GenAI tools appeal to L2 writing teachers and students due to such adaptive feed-

back to student compositions as suggestion, explanation, and even full-text modeling. Much of the existing literature in this realm advocates the integration of this affordance into L2 writing pedagogical design in settings of individual writing (Muñoz Muñoz et al., 2025; Ou et al., 2024; Chen & Lv, 2024; Su et al., 2023; Wambsganss et al., 2022; Wei & Li, 2023), collaborative writing (Wiboolyasarini et al., 2024), and student feedback training (Guo et al., 2024; Rad et al., 2023). The prevalence of genAI-assisted writing has yielded the view that students' ability to leverage genAI feedback in writing activity, including sending effective prompts, diagnosing the accuracy of genAI feedback, and incorporating it appropriately and ethically, is an indispensable part of their AI literacy (Barrios-Beltran, 2025; Praphan & Praphan, 2023; Warschauer et al., 2023).

Student engagement with this affordance is essential to feedback leveraging results and writing learning. However, very limited empirical study has comprehensively examined multi-faceted student engagement with genAI feedback in L2 writing activities. Therefore, our study unfolds how student writers respond to genAI feedback affectively, cognitively, and behaviorally in a Chinese tertiary English as a foreign language (EFL) course. Additionally, some pedagogical proposals are advanced to help teachers optimize their genAI feedback-enhanced writing pedagogy design.

2. Literature Review

2.1. Research on genAI Feedback in L2 Writing

Utilizing genAI tools for writing feedback through human-machine communication is underpinned by sociocultural theory (Vygotsky, 1978), interactive theory (Long, 1996), and constructivist learning theory (Wegerif, 2004), which all emphasize learners' knowledge construction through interaction. Previous literature on genAI feedback in L2 writing field mainly falls into three strands: the feasibility of integrating genAI feedback into L2 writing pedagogy, teacher and student perceptions of genAI as a feedback source, and the effects of using genAI feedback on students and their writing.

Most academics in L2 writing field have acknowledged the legitimacy of harnessing genAI feedback in writing pedagogy. The most frequently listed advantages of the feedback are catering to personalized requirements, alleviating writing anxiety, and reducing writing teachers' repetitive evaluating workload (Barrios-Beltran, 2025; Barrot, 2023; Chen & Lv, 2024; Liu et al., 2024; Mizumoto & Eguchi, 2023; Su et al., 2023; Tseng & Warschauer, 2023; Warschauer et al., 2023; Wei & Li, 2023). For example, genAI feedback can help to optimize writing quality at macro and micro levels, including text clarity, depth of analysis, style, consistency, grammar, vocabulary, etc. (Barrot, 2023; Wei & Li, 2023). It can also present revision schemes progressively, from general assessment, through corrective recommendation, to metalinguistic explanation, and finally to revised version provision (Chen & Lv, 2024). An additional reason for its popularity among students is their access to genAI feedback "without any fear of being judged" (Barrot,

2023; p. 2). Admittedly, some scholars (Wei & Li, 2023; Liu et al., 2024) expressed concerns about students' losing independent thinking due to overreliance on genAI models. For writing teachers, genAI feedback can release their heavy composition evaluating work besides scoring particularly in large-class contexts (Mizumoto & Eguchi, 2023).

The second strand of scholars in this field delves into L2 teacher and student perception of genAI models as a feedback source. Teachers are judicious about the role of such models as evaluators (Abendschein et al., 2024; Escalante et al., 2023; Guo & Wang, 2023; Ou et al., 2024). For instance, ChatGPT is viewed as an evaluator motivating learners to write and correct form errors, yet it is also criticized for not detecting off-topic points and constraining low L2-proficiency learners from reading lengthy comments (Guo & Wang, 2023). Like teachers, students display mixed attitudes towards these feedback providers (Abendschein et al., 2024; Cummings et al., 2024; p. 9; Escalante et al., 2023; Ou et al., 2024; Ruwe & Mayweg-Paus, 2023). They regard genAI models as competent trustworthy teachers (Abendschein et al., 2024; Ou et al., 2024) or even more reliable than teachers due to the models' massive L2 writing repertoire, therefore highly acknowledging genAI feedback (Ruwe & Mayweg-Paus, 2023). However, they worry that those tools deprive of teacher-student connection in writing process, which impairs their writing learning (Cummings et al., 2024).

The third strand probes into the effects of genAI feedback on student writers and their L2 writing proficiency (Annamalai et al., 2023; Meyer et al., 2024; Muñoz Muñoz et al., 2025; Ou et al., 2024; Ruwe & Mayweg-Paus, 2023; Wambsganss et al., 2022; Wiboolyasarin et al., 2024; Zhang & Hyland, 2025). Firstly, mixed conclusions have been reached on the effects of genAI feedback on students. For example, Ruwe and Mayweg-Paus's (2023) research revealed that genAI's personalized feedback language did not enhance students' writing self-efficacy and overall writing motivation. However, Meyer et al. (2024) reported that genAI assessment boosted student motivation in addressing similar writing tasks. Similarly, Ou et al. (2024) discovered that genAI feedback helped to reinforce students' agency and authority during L2 writing. Secondly, contradictory results have been obtained regarding the effects of genAI feedback on L2 learners' writing proficiency. For example, Muñoz Muñoz et al. (2025) found that ChatGPT-generated direct corrective written feedback was superior to the teacher's feedback in improving student writing regarding task response, cohesion and coherence, lexical resource, and grammatical range and accuracy. Although in Meyer et al.'s (2024) study, genAI feedback was also proved to facilitate participants' revision, it was not a decisive factor for their improvement in subsequent writing. However, Wiboolyasarin et al.'s (2024) investigation verified that ChatGPT feedback contributed to students' substantial writing improvement in content, organization, and language.

Enlightening as these researches on genAI feedback in L2 writing are, what remains to be unfolded is how student writers as the core agents engaging with such

feedback, an important intermediary link that decides the effect of feedback (Ellis, 2010; Zhang, 2020) in genAI-enhanced writing context.

2.2. Construct of Student Engagement with Feedback

Ellis (2010) defines student engagement as “how learners respond to the feedback they receive” (p.342). Drawing on Fredricks et al.’s (2004) framework of school engagement, Ellis (2010) proposes the tripartite conceptualization of student engagement with feedback that encompasses three interdependent components: affect, cognition, and behavior. According to Ellis (2010), affective engagement refers to students’ attitude to feedback; cognitive engagement the extent to which students attend to feedback; behavioral engagement the way students uptake feedback.

This conceptualization lays a foundation for extensive investigation of student engagement with teacher feedback, peer feedback, and automated writing evaluation (AWE) feedback (Han & Hyland, 2015; Fan & Xu, 2020; Koltovskaia, 2020; Tian & Zhou, 2020; Zhang & Hyland, 2018; Zheng & Yu, 2018). Affective engagement is commonly operationalized as students’ interest in feedback, perception of its value, and affect after receiving it (Fan & Xu, 2020; Han & Hyland, 2015). Cognitive engagement is manifested as feedback noticing, the degree of understanding it, cognitive and metacognitive operations deployed in mentally processing it (Fan & Xu, 2020; Han & Hyland, 2015; Koltovskaia, 2020). Reported cognitive operations include such strategies as evaluating the correctness of feedback, orally repeating feedback, taking notes, retrieving previous knowledge, and deciding revision. Metacognitive operations are measured by students’ regulation of their mental effort such as planning and monitoring. Behavioral engagement is manipulated mainly with three indicators: revision time (Koltovskaia, 2020; Zhang & Hyland, 2018); revision operations, i.e., ways of feedback utilization (Fan & Xu, 2020; Han & Hyland, 2015; Zheng & Yu, 2018; Koltovskaia, 2020); and revision strategies, i.e., consulting external sources of teachers, peers, and internet (Fan & Xu, 2020; Han & Hyland, 2015; Koltovskaia, 2020).

Predicated on the multi-dimensional framework of student engagement with human and AWE feedback, the present study also treats student engagement with genAI feedback as a meta-construct composed by the interrelated dimensions of affect, cognition, and behavior. Specifically, affective engagement relates to students’ attitude towards genAI feedback; cognitive engagement concerns their mental efforts in processing genAI feedback and deciding revisions; behavioral engagement concerns revision practice.

2.3. Empirical Research on Student Engagement with genAI Feedback in L2 Writing

Empirical studies on student engagement with genAI feedback mostly focus on affective dimension (Abendschein et al., 2024; Barrios-Beltran, 2025; Escalante et al., 2023; Ou et al., 2024). Learners are generally positive towards genAI feedback

for its clarity, specificity, and credibility. For example, Escalante et al. (2023) found participants appreciated GPT-4's feedback in improving their academic writing for its easy availability, concrete explanation of errors, and suggested solutions. Learners' disappointment with genAI feedback are also documented. They complain about the incorrectness of genAI feedback especially in non-EFL writing contexts, and worry that feedback in the form of full-length model texts is detrimental to writing learning, independent thinking, and academic integrity. For example, in Ou et al.'s (2024) investigation of ChatGPT-assisted academic communication, the students were annoyed by repetitive wrong feedback to their academic formulation errors.

Koltovskaia et al.'s (2024) multi-faceted study of graduate students' engagement with ChatGPT feedback is quite enlightening. In their English research proposal writing, the participants affectively appreciated the paraphrasing functionality of this chatbot for more professional writing, cognitively gauged the accuracy of feedback, and behaviorally revised lower-order errors.

Despite the enlightenments from previous literature about student involvement with genAI feedback, the review indicates the urgency of more studies to probe this involvement from the tripartite lens by adding cognitive and behavioral dimensions for a more complete picture of this construct.

3. The Current Study

Underpinned by Ellis' (2010) multi-dimensional conceptualization of student engagement, our study aims to bridge the gap of less integrated investigation by scrutinizing how EFL students engage affectively, cognitively, and behaviorally with the feedback from Kimi (a genAI chatbot) in a college English course. To be specific, this research attempts to answer the following questions:

- 1) What is the general pattern of Kimi feedback prompted by EFL students?
- 2) How do EFL students engage with Kimi feedback affectively, cognitively, and behaviorally?

4. Methods

4.1. Context and Participants

The study was undertaken in a key Chinese public polytechnic university. A class of 53 students (aging 20 and 21 years old) were recruited in the study when they were at the beginning of the first-year undergraduate study. Their English proficiency level was upper-intermediate. Enrolled in a college English course, they met twice a week for a 90-min session each time over 12 weeks. One pedagogical goal of the course syllabus is to develop students' skills of writing English argumentative essay. Most participants (71.7%) had no experience of using genAI feedback in high school English writing learning. Informed consent about the purpose of the research and the use of the data was obtained from all participants. Their names in the study are pseudonyms.

4.2. Kimi

Kimi is an online (<https://kimi.moonshot.cn/>) genAI chatbot capable of processing and generating natural language. It can provide real-time affordances such as forming writing outline, scoring texts, and generating model texts. An appealing affordance is providing adaptive feedback on texts concerning thesis development, text unity, idea organization, grammar, vocabulary, etc. This functionality allows students extensive opportunities to learn English writing. Moreover, the process of student-genAI interaction can be preserved as a web link for post-writing reflection and pedagogical analysis of writing process.

4.3. Data Collection

The data was collected throughout a period of 8 weeks (see **Table 1**). In week 1, participants acquired genre knowledge of English argumentative essay via teacher-recorded pedagogical videos (step 1) and received prompt training for Kimi feedback (step 2). In this training session, after introducing Kimi’s feedback functionality, the teacher (the first researcher) demonstrated prompting with a previous student’s argumentative essay. Those prompts pertained to content (clarity, unity, etc.), organization (coherence and cohesion), and form (grammatical, lexical, and mechanical issues). The teacher also showed human-Kimi “discussion” (e.g., re-prompting for more precise feedback or consulting for explanation of mentioned writing knowledge). Meanwhile, the teacher emphasized the necessity of critical thinking due to potential inappropriate Kimi feedback. Afterwards, the students practiced prompting with an argumentative essay by another previous student. To facilitate student-genAI discussion, two supports were available: an argumentative essay-scoring rubric (see **Appendix A**) adapted from the assessing criteria by Guo, et al. (2024), Chen (2019), and Langan (2014: p. 151); a teacher-developed prompt handout (**Appendix B**). Other sources such as dictionaries, classmates, and the teacher were recommended for students to confirm Kimi feedback.

Table 1. Procedures for data collection.

Time	Procedure	Data collected
Week 1	Step 1: introduction to argumentative essay	
	Step 2: prompt training	
Week 2	Step 3: the first after-class argumentative essay	
Week 8	Step 4: the final after-class argumentative essay	First drafts, final drafts, archived Kimi records
	Step 5: assigning a reflection paper	Reflection papers
	Step 6: questionnaire survey	Questionnaire data
	Step 7: one-on-one semi-structured interviews	Audio recordings of interviews

In week 2, the first after-class argumentative essay was assigned (step 3). The

students consulted Kimi for feedback to direct their revision of the first draft. From week 3 to week 7, they used this feedback affordance in following argumentative writing practice.

In week 8, the final after-class argumentative essay on the integration of AI in higher education was assigned (step 4). The participants revised the first draft guided by Kimi feedback and wrote a reflection paper about their cognitive investment (see **Appendix C** for prompt questions; step 5). The score for each final draft was reached by averaging the scores from the first researcher and Kimi based on the same scoring rubric provided to the participants with a scale of 0 - 20 points. Then, the participants were classified into three groups accordingly: high group (H) receiving 20 - 17 points, middle group (M) 16 - 13 points, and low group (L) less than 13 points. Upon the completion of the reflection paper, a questionnaire about affective engagement (see **Appendix D**) adapted from [Fan and Xu \(2020\)](#) and [Tsui and Ng \(2000\)](#) was immediately administered via Wenjuanxing, a free online survey applet in China (step 6). Three affective subdimensions are indicated by 7 items, with items 1 - 2 indicating the interest and willingness in receiving Kimi feedback, items 3 - 4 feelings after receiving feedback, and items 5 - 7 the value of Kimi feedback. The original questionnaire has demonstrated good content validity in previous studies. Additionally, we invited two experts in L2 writing to verify the appropriate alignment of each item with the target concept. Participants rated each item on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

To unveil student engagement with Kimi feedback more thoroughly, individual semi-structured interviews (see **Appendix E** for guiding questions) were conducted within two days of final draft submission (step 7) with three representatives randomly selected from each group (see **Table 2** for profiles). The interviews were conducted in Chinese in the teacher's office and recorded. The first and final drafts were used as stimuli to recall writing experience with Kimi feedback.

Table 2. Profiles of interviewees.

Name	Gender	Major	Group
Hao	Male	Geomatics engineering	H
Long	Male	Geomatics engineering	H
Juan	Female	Mathematics	H
Yi	Female	Geomatics engineering	M
Chen	Male	Electrical engineering	M
Hong	Female	Mathematics	M
Qin	Male	Geomatics engineering	L
Yang	Male	Geomatics engineering	L
Han	Male	Mathematics	L

Note: H = high group; M = middle group; L = low group.

Through the above procedures, research credibility was realized via data triangulation with six sources of data: archived Kimi records, questionnaire, reflection papers, first drafts, final drafts, and audio recordings of interviews. The first researcher transcribed the audio recordings of interviews and translated the transcripts as well as reflection papers into English. The second researcher checked the accuracy of the transcripts and translation data against original materials.

4.4. Data Analysis

The data analysis consisted of two parts: descriptive analysis of feedback in archived Kimi records, revisions in final drafts, as well as questionnaire data; content analysis of reflection papers, interview transcripts, and students' consultation in archived Kimi records.

4.4.1. Analysis of Feedback in Archived Kimi Records and Multiple Drafts

Kimi feedback was analyzed to disclose its pattern in terms of distribution and accuracy. All feedback points in archived Kimi records were documented based on Hyland's (1998) definition of feedback point, i.e., "a written intervention that focused on a different aspect of the text" (p. 261). They were then coded as content-focused, organization-focused, and form-focused ones. Content-focused feedback refers to comments on argument clarity, argument development with varied details for persuasiveness, and effective counterargument addressing (Guo et al., 2024). Organization-focused feedback involves evaluations on essay structure, including coherence and transition between ideas and paragraphs (Langan, 2014). Form-focused feedback relates to comments on lexical, grammatical, and mechanical errors (Ashwell, 2000; Han & Hyland, 2015). The accuracy of Kimi feedback was also checked through comparison of the feedback points with first drafts. The criteria for accuracy are a comment 1) optimizes the original meaning, 2) clarifies the structural logic, or 3) corrects form-related error while aligning with the intended meaning. For example, S38 wrote "Additionally, although AI technology helps to promote the sharing of high-quality resources, in practice, some high-quality AI educational resources may be concentrated in developed areas or high-income families, which could further exacerbate the imbalance in educational resource distribution and affect the fairness of education". Kimi evaluated "This sentence is grammatically correct, but it could be simplified for clarity" and corrected it as "Additionally, while AI technology promotes the sharing of high-quality resources, some of these resources may be concentrated in developed areas or high-income families, potentially exacerbating educational inequality". This feedback point was accounted as accurate due to its successful recognition of redundant expression. The following feedback point was categorized as inaccuracy. Kimi misjudged "for" in "Education is highly concerned for its shaping human beings' development" (S42) as an error and explained that "'is highly concerned with' should be used to express the idea of being focused on a particular aspect". However, the context of the sentence in the first draft revealed that "shaping human beings' development" is the reason for the importance of education. Such

logic should be realized by adding “with” before “for” rather than Kimi’s inappropriate suggestion of substituting “for”. Thus, each category of feedback was further divided into accurate and inaccurate subcategories. To ensure coding consistency, the second researcher randomly coded 12 participants’ (around 23% of the class) archived Kimi records and first drafts. The inter-coder reliability for feedback categories was 95%. Discrepancy was resolved through discussion. Finally, we tallied the number of each (sub) category of feedback and calculated the percentages. In this study, general admonitions (Yu et al., 2019) of essays did not emerge; emotional feedback was excluded for it does not lead to participants’ revision behavior (Hyland, 1998).

Participants’ multiple drafts and archived Kimi records were then examined to identify revision operations as the reflection of behavioral engagement. The first researcher located Kimi-triggered revisions by comparing the differences between first and final drafts and then verifying whether there were corresponding comments for the changes in archived Kimi records. The cases where no revision occurred with Kimi feedback were also documented. Drawing on Koltovskaia (2020) taxonomy of revision operations as acceptance and rejection, all revision operations in our study were coded as four types: acceptance of accurate feedback, rejection of accurate feedback, acceptance of inaccurate feedback, and rejection of inaccurate feedback. The second researcher still coded the 12 participants’ drafts and archived Kimi records. The result of inter-coder reliability check for revision operations was 97%, with discrepancy resolved through discussion. Similarly, the number of each (sub)category of revision was tallied for percentage calculation.

4.4.2. Analysis of the Questionnaire

The questionnaire data ($n = 53$) were examined through SPSS 29 for mean values, standard deviations of each item to uncover the participants’ affective engagement.

4.4.3. Analysis of Reflection Papers, Consultation in Archived Kimi Records, and Interviews

Content analysis (Krippendorff, 2004) was employed to examine three sources of collected data: reflection papers and students’ consultation in archived Kimi records as the evidence for cognitive dimension; interview as complementary information of the three dimensions. These data sets were iteratively read and examined to identify segments reflecting affective, cognitive, and behavioral engagement. In Vivo Coding (Miles et al., 2020) was used on the segments to accurately capture participants’ engaging experience. Those preliminary codes were then compared constantly and clustered into categories based on their similarities or meaningful relevance, which were revised and refined before being further merged into themes. Consequently, the construct of student engagement with Kimi feedback was conceptualized as shown in Table 3. To ensure the reliability of data analysis, the three data sources of the above 12 participants were coded by the

second researcher. The inter-coder agreement reached 93%, with disagreements being resolved by discussion.

Table 3. Conceptualization of student engagement with Kimi feedback.

Dimension	Subdimension	In Vivo Code	Example
Affective engagement	Overall interest in Kimi feedback	“Efficient”	“It is really efficient to help me identify the mistakes that I am unaware of.” (Yi, Interview)
	Willingness to interact with Kimi	“Re-prompted”	“In the beginning, the feedback was too general. So, I re-prompted it to ask which part needed to be revised.” (Qin, Interview)
	Evaluation of the worth of Kimi feedback	“Point out”	“It can point out my contradictory views.” (Yi, Interview)
	Feelings after receiving Kimi feedback	“Happy”	“I was happy to see so many feedback from Kimi as if there was a person helping me improve my writing.” (Yang, Interview)
Cognitive engagement	Understanding of Kimi feedback	“Understand”	“I can understand Kimi feedback easily.” (Huang, Interview)
	Cognitive operations	“Read”	“I read the feedback carefully and then decided addition or deletion.” (S8, Reflection paper)
	Metacognitive operations	“Evaluate”	“I read the whole essay to evaluate whether the revisions within paragraphs or sentences disrupted the logic.” (S43, Reflection paper)
Behavioral engagement	Revision operations	“Took it up”	“Kimi told me the word ‘massive’ was too big and ‘significant’ was more exact. I agreed that Kimi was right and I took it up .” (Long, Reflection paper)

5. Results

5.1. General Pattern of Kimi Feedback

Kimi feedback to the participants’ first drafts of the final argumentative essay shown archived Kimi records were analyzed. The result revealed the pattern of this student-prompted feedback, as presented in **Table 4**. In terms of focus distribution, among all feedback (535), form-focused feedback (351, 65.61%) were much more than organization-focused (102, 19.07%) and content-focused (82, 15.33%) ones. Regarding accuracy, the total feedback accuracy reached 86.73% (464). The rates of accurate form-focused, organization-focused, and content-focused feedback were 87.75% (308), 85.29% (87), and 84.15% (69), respectively. Within the accurate content-focused feedback, more were related to argument development (36, 52.17%) and counterargument addressing (21, 30.43%). The accurate organization-focused comments concentrated more on logical and integrated body paragraphs (30, 34.48%) and use of transitions (31, 35.63%). The accurate form-related comments converged more on lexical (164, 53.25%) and grammatical errors (110, 35.71%). It is worthy of noting that 71 (13.27%) comments were

inaccurate, mainly pertaining to lexical (28, 39.44%) and grammatical errors (13, 18.31%).

Table 4. Kimi feedback pattern and revision operations.

Type	Subtype	AF	A	R	IF	A	R
CFF (82)	Argument clarity	12	5	7	3	1	2
	Argument development	36	30	6	4	0	4
	Counterargument addressing	21	10	11	6	1	5
	Subtotal	69	45	24	13	2	11
OFF (102)	Clear and engaging introduction	7	6	1	1	0	1
	Logical and integrated body paragraphs	30	17	13	4	0	4
	Use of transitions	31	24	7	6	0	6
	Powerful conclusion	19	11	8	4	0	4
	Subtotal	87	58	29	15	0	15
FFF (351)	Lexical errors	164	141	23	28	6	22
	Grammatical errors	110	98	12	13	7	6
	Mechanical errors	34	27	7	2	0	2
	Subtotal	308	266	42	43	13	30
Total		464	369	95	71	15	56

Note: CFF = content-focused feedback; OFF = organization-focused feedback; FFF = form-focused feedback; AF = accurate feedback; IF = inaccurate feedback; A = acceptance; R = rejection.

5.2. Affective Engagement

The analysis of the questionnaire data and interviews revealed the participants' high affective engagement with Kimi feedback. As is shown in **Table 5**, they were very positive to Kimi feedback across a range of questionnaire items, with mean values ranging from 1.45 (SD = 0.748) on embarrassment after receiving Kimi feedback to 4.57 (SD = 0.694) on its perceived worth in improving essay form.

Comments from the interviewees confirmed this high embracement of Kimi feedback. It was described as “convenient”, “immediate”, “easily accessible”, “efficient”, and “extensive”. This positive attitude was best summarized in Hao's words:

“With abundant database, it can indeed provide many benefits to my essay revision. It is very convenient. If I want suggestions on certain aspects, I just need to input related prompts to Kimi. I will use genAI feedback in other English writing tasks even if the teacher doesn't require me to do so.” (Interview)

Table 5. Descriptive statistics for affective engagement (n = 53).

No.	Item	M	SD
1	Interest in Kimi feedback	4.26	0.812
2	Willingness to negotiate with Kimi	4.17	0.893
3	Happiness after receiving Kimi feedback	3.98	0.930
4	Embarrassment after receiving Kimi feedback	1.45	0.748
5	Helpfulness in content	4.34	0.807
6	Helpfulness in organization	4.19	0.878
7	Helpfulness in form	4.57	0.694

They also exhibited strong interest in negotiation with Kimi. Unconstrained inquiry to Kimi allowed them to probe writing problems sufficiently and expand writing knowledge. Besides, some respondents acknowledged that this communication with Kimi about its possible inaccurate feedback could force them to “discern right from wrong”, thus enhance their “critical thinking” (Chen, Interview).

They also expressed the happiness after receiving Kimi feedback. They were delighted when the errors in the drafts were pointed out because they knew the purpose of using genAI feedback was to know the drawbacks of an essay and solve problems. However, some of them were dissatisfied with its lengthiness. This was reflected in Hao’s words:

“Even if I had prompted Kimi to reduce the number of feedback words, it neglected my requirement and produced long feedback most of time. I hadn’t so much time to read all these comments carefully. Also, such long feedback makes me doubt my writing ability.” (Interview)

Generally, learners appreciated the value of Kimi feedback in improving writing quality. For example, Hao regarded most feedback as “pertinent and very useful” for his revision, conducive to enhancing his English writing level (Interview). Yi acknowledged its ability to point out her “contradictory ideas” (Interview). Particularly, learners acknowledged its assistance in refining word choice, correcting grammatical and spelling errors, and reinforcing the use of connective words. As Long noted:

“It can rapidly provide detailed explanation on my errors of sentence patterns and spelling. Particularly, I tend to make spelling errors. I often fail to recognize those errors, but Kimi can mark them.” (Interview)

Meanwhile, a few respondents doubted the worth of Kimi feedback for its overgenerality and inappropriateness. For example, Yi mentioned that Kimi feedback was “not so effective in solving theme-related problems because it was not so targeted” (Interview). Juan complained that Kimi as a tool sometimes could not capture her intention, thus generating useless suggestions.

5.3. Cognitive Engagement

Data analysis of reflection papers, archived Kimi records, and interview indicated participants' overall high cognitive investment in Kimi feedback. Most of them (40/53) reported complete understanding of Kimi feedback owing to its clear language and detailed explanation. Nine participants understood most of it and four partially mainly due to writing terms, especially grammatical terms in feedback.

Various cognitive operations were deployed in processing Kimi feedback and deciding revision operations (see **Table 6**). The most frequently mentioned device was comparing Kimi feedback with original writing (28/53). Participants scrutinized whether the suggestions conformed to the intended meaning, improved essay quality, or impaired the consistency of the essay. For example, S8 stated he would “primarily checked whether Kimi understood my meaning” to “ensure the suggestion fully fit into the essay context before adopting it” (Reflection paper). Han chose to accept the suggestions that were “appropriated for my topic” and “made my essay academic” (Interview). Hao treated the matching of feedback with the original meaning as a principle to accept a suggestion, thus filtering out “those inappropriate, irrelevant, or repetitive ones” (Interview).

Table 6. Cognitive operations in response to Kimi feedback (n = 53).

Cognitive operation	Number of participants
Comparing Kimi feedback with original writing	28
Applying English proficiency and genre-specific knowledge	22
Leverage Kimi for writing knowledge	17
Leverage other software for translation, writing knowledge, or feedback	15
Reiterative reading of Kimi feedback	15

Applying English proficiency and genre-specific knowledge was the second most frequently used cognitive operation (22/53). Participants capitalized on their grammatical knowledge, lexical knowledge, and argumentative discourse knowledge to judge the accuracy of Kimi feedback prior to revision. For example, S12 denied Kimi's suggestion of using more advanced vocabulary because “overuse of advanced words would make the article obscure” (Reflection paper). Yi dismissed Kimi's recommendation of including “a specific prediction or a call to action to motivate readers” based on her genre knowledge of “multiple ways of ending an argumentative essay rather than just a prediction or an appealing” (Interview). Juan accepted Kimi's suggestion of adding explicit connectives between discussed points by drawing on her coherence knowledge.

Participants also inquired Kimi or other software for translation and writing knowledge, or verified the appropriateness of Kimi feedback with comments from

other software. For example, Kimi was consulted for transitional expressions in academic writing, further explanation of its suggestions, and listing ways of conveying or retorting a counterargument and introducing a compelling anecdote. This is how Yi decided her acceptance of Kimi feedback:

“I could not understand why Kimi suggested me to replace ‘change in’ with ‘change to’, so I prompted it to explain. It told me that the preposition after ‘change’ relies on the phrasal verb I use and my intended meaning. I should choose ‘to’ because I used ‘brought’ before ‘change’.” (Reflection paper)

Another cognitive operation was reiterative reading of Kimi feedback. Yang, a representative participant, depicted re-reading Kimi feedback as “an important thinking process” even though he “might reject some feedback” (Interview).

However, some participants invested little cognition in processing certain Kimi feedback, which led to their blind acceptance of inappropriate recommendations. This is intensively reflected in Hong’s words:

“Some suggestions made me confused. For example, Kimi suggested me to change ‘which’ in the sentence ‘AI also fosters laziness in our study and makes us lack independent thinking, which is essential in our studies’ into ‘that’. I could not understand, but I revised it because I thought Kimi was usually correct.” (Interview)

The use of metacognitive operations was reported by most participants (40, 75.47%) in reflection papers and interview. The deployed operations included planning, monitoring, and reiterative reading the revised parts or even the whole draft to evaluating the appropriateness of revisions (see **Table 7**).

Table 7. Metacognitive operations in response to Kimi feedback (n = 53).

Metacognitive operation	Number of participants
Reiterative reading the revised parts or even the whole draft to evaluating the appropriateness of revisions	24
Planning	18
Monitoring	5

The most frequently reported metacognitive strategy was reiterative reading the revised parts or even the whole draft to ensure appropriately revised essays (24/53). For instance, S43 mentioned that he “re-examined the organization of the essay to prevent revisions within paragraphs and sentences from weakening the wholistic logic of the essay” (Reflection paper).

Another frequently reported metacognitive strategy was planning (18/53). Specifically, two types of planning in processing Kimi feedback were found: from higher-order to lower-order issues or vice versa. Some participants addressed content- and organization-focused concerns first because they viewed these issues

“mattered most” (S1, S12, Reflection paper). As Juan noted:

“I checked the feedback on organization first. After solving these key problems, I moved to the feedback on sentence patterns, and finally on vocabulary. This made my revision more efficiently.” (Interview)

The participants prioritizing form-focused feedback reported their tendency to make grammatical, lexical, or spelling mistakes in English writing. Therefore, they preferred to “solve those prominent issues with priority” (S12, Reflection paper). Monitoring was also adopted to regulate revision process (5/53). Some participants checked “if any suggestion was missed” (S12, Reflection paper) or reflected “if too much attention” was on “vocabulary and grammar” (S27, Reflection paper).

5.4. Behavioral Engagement

Participants’ behavioral investment in Kimi feedback was manifested as their revision operations, i.e., acceptance or rejection of Kimi feedback in final drafts. These operations were confirmed by archived Kimi records, reflection papers, and interviews after the final writing task. As is shown in **Table 4**, the total adoption of accurate feedback was high (369, 79.52%). Among them, the adoption of accurate form-focused feedback (266, 86.36%) was higher than those of accurate content-focused (45, 65.22%) and organization-focused (58, 66.67%) feedback. Within accurate content-focused feedback, more adoption was about argument development (30, 83.33%). Accurate organization-focused feedback was highly accepted across its four sub-categories: clear and engaging introduction (6, 85.71%), logical and integrated body paragraphs (17, 56.67%), use of transitions (24, 77.42%), and powerful conclusion (11, 57.89%). Similar high acceptance was observed on accurate form-related comments across its sub-categories: lexical errors (141, 85.98%), grammatical errors (98, 89.09%), and mechanical errors (27, 79.41%). Additionally, the participants were capable of dismissing most inaccurate feedback (56, 78.87%). The rejection rates of inaccurate content-focused, organization-focused, and form-focused feedback were 84.62% (11), 100% (15), and 69.77% (30), respectively.

Such high adoption of the accurate Kimi feedback and rejection of inaccurate feedback could probably be attributed to their cognitive and metacognitive efforts that facilitated their understanding and scrutinizing of genAI suggestions. Meanwhile, many participants mentioned the motivation to write “a high-quality essay for a high score” (S38, Reflection paper). This might translate into their positive behavioral investment, especially in effort-demanding higher-order issues.

Naturally, some participants with insufficient cognitive investment and low motivation refused reasonable but less specific content- and organization-focused feedback. As is reflected in the case of Hong:

“Feedback on logic and organization are not as specific as that on form. I needed to figure out specific revisions, so I refused most of them. For exam-

ple, I knew the logic in my second paragraph was not clear, but Kimi just told me to strengthen it without giving me specific advice. I didn't know how to revise. I am lazy. I didn't take it." (Interview)

6. Discussion

6.1. Kimi Feedback Pattern

Our results about Kimi feedback pattern in terms of its multiple levels and high accuracy contribute to the line of research that supports the integration of genAI models in EFL writing. Firstly, Kimi in our study yielded comments pertaining to form, content, and organization (e.g., argument development and coherence) as elicited by participants. This multi-level feedback pattern could partially be attributed to the guidance of teacher-provided prompt handout. However, as the participants were just suggested to use it, the pattern corroborates the finding of previous literature that EFL student writers pay attention to higher- and lower-order issues (Fan & Xu, 2020, Yang & Xu, 2022). Meanwhile, this pattern proves genAI models such as Kimi have the capacity to accommodate their extensive needs by providing multi-leveled suggestions (Barrot, 2023; Wei & Li, 2023).

Secondly, our study showed Kimi's capacity of precisely detecting and analyzing students' different levels of writing errors (i.e., form, content, and organization). This observed reliability of Kimi feedback differs from Koltovskaia et al.'s (2024) finding about ChatGPT's 50% of accuracy in identifying and commenting lower-order issues. One reason for the divergence might be students' different human-AI interactive negotiation competence (Wen & Liang, 2024) in eliciting effective genAI feedback. In our study, participants experienced 6 weeks of English writing assisted by Kimi as the feedback provider. This possibly enhanced their human-AI interactive negotiation competence in prompting Kimi to generate accurate feedback on both local and global aspects. However, learners in Koltovskaia et al.'s (2024) study might be less skilled in sending precise directions just after the in-class genAI-based writing training.

A pedagogical implication of our finding is that to acquire accurate genAI feedback in writing activity, EFL learners need to develop communicating competence in fine-tuning prompts regarding content, organization, and form. A prompt handbook as reference is necessary for students to use appropriate lexical, syntactical, and discourse terms. Also, repetitive training on sending multiple levels of prompts is encouraged to reinforce EFL student writers' awareness of higher-order issues.

6.2. EFL learners' Affective, Cognitive, and Behavioral Engagement

This study extends the previous research on student engagement with genAI feedback by expounding EFL learners' cognitive and behavioral response besides affective response to Kimi feedback. It provides empirical evidence of EFL learners' high involvement of genAI feedback in writing activity.

Affectively, learners in our study generally demonstrated positive attitudes to-

wards Kimi feedback, which resonates with previous findings about students' embracement of genAI for text revision (Abendschein et al., 2024; Escalante et al., 2023; Koltovskaia et al., 2024). Specifically, our participants displayed contentment with its convenience, easy access, efficiency, and wide coverage of different levels of writing errors. They were also willing to communicate with the model, perceiving this interaction as the opportunity to expand writing knowledge and enhance critical thinking. After receiving Kimi feedback, they appreciated Kimi for enabling them to know writing drawbacks although some of them complained the time-consumption of reading long comments. Few participants felt the embarrassment of seeing their writing products corrected, which, however, was reported in Fan and Xu's (2020) research on student engagement with peer feedback. A reason for the difference might be our reflexing human-AI interaction setting where no psychological barrier of being judged by peers existed (Barrot, 2023). Besides, the present study showed participants' general recognition of Kimi's value in improving writing quality from higher- to lower-order aspects despite some learners' dissatisfaction with Kimi's vague content suggestion. Such positive judgment of Kimi's value in writing revision could be explained by the high accuracy of Kimi feedback owing to participants' enhanced competence in human-AI communication (Wen & Liang, 2024). This general positive affective engagement may help learners mobilize their cognitive and behavioral involvement.

Cognitively, mental efforts were observed in learner's processing Kimi feedback and deciding its incorporation. Most participants reported full understanding of Kimi comments. They deployed various cognitive operations: comparing Kimi feedback with original writing; applying their English proficiency and genre-specific knowledge; leveraging Kimi or other software for translation, writing knowledge, or feedback; reiterative reading of Kimi feedback. These operations, defined as a kind of AI literacy in Praphan and Praphan (2023) and Warschauer et al. (2023), helped participants discern Kimi feedback in mixed quality and decide adoption. It is worth noticing that some learners exhibited outstanding AI literacy by drawing on various cognitive operations. Some participants, however, exhibited inadequate AI literacy due to excessive trust in genAI models, which led to their blind acceptance or neglect of Kimi feedback on higher- or lower-order errors. This result supports the warning of Wang and Wang (2024) that large language models may impede EFL learners' motivation for deep knowledge comprehension.

Many participants also leveraged metacognitive operations to regulate mental activity throughout revision. They developed feedback processing sequence according to their primary concerns (i.e., content, organization, or form) and reiteratively read the revised parts or essays to evaluate the appropriateness of revisions. A few participants monitored their revision to ensure attention to all feedback or proper time allocation to each type of feedback. Such metacognition was also reported in Koltovskaia et al.'s (2024) study with graduate students but rarely in Zheng and Yu's (2018) investigation of lower-proficiency EFL undergraduates. Thus, English language proficiency might be a predictor of metacognitive effort

(Sun et al., 2023). The upper-intermediate English level of our participants enabled them to leverage previous writing experiences for revision judgment and task management. However, a few cases in our study deployed no metacognitive operations. This indicates that deficient critical thinking awareness possibly impedes students even at the upper-intermediate English proficiency level from exerting metacognitive operations.

The pedagogical implication of our learners' complexity in harnessing cognitive and metacognitive operations is the significance of enhancing student writers' critical thinking in genAI feedback-facilitated EFL writing. Such AI literacy including critical thinking awareness and ability is decisive in judging the legitimacy of genAI feedback, deciding feedback adoption, and evaluating the quality of feedback-based revision. To avoid excessive reliance on genAI tools in EFL writing, genAI feedback session could be followed by peer discussion or teacher-student conference. Thus, peer or teacher feedback could be used as verification of or complement to genAI feedback.

Behaviorally, the current study showed participants' high adoption of accurate feedback and rejection of inaccurate feedback on content, organization, and form. In terms of accurate Kimi feedback, they incorporated the suggestions related to argument development, text coherence, persuasiveness of conclusion, lexical errors, grammatical errors, etc. Our learners' high acceptance of the three types of Kimi feedback is partially contrary to the findings of previous literature on peer, AWE, and genAI feedback that showed learners' extensive uptake of lower-order suggestions but much less of higher-order ones (Fan & Xu, 2020; Koltovskaia et al., 2024; Zhang, 2020). Additionally, our participants rejected most inaccurate Kimi feedback on both higher- and lower-order issues. Similar dismissal was rarely reported in other researches with genAI tools except in Koltovskaia et al.'s (2024) case study. A few factors may explain our students' positive behavioral investment in accurate and inaccurate feedback on various errors. One possible factor is the high accuracy of feedback especially on higher-order issues. This implied increased credibility and usability, thus provoking learners' positive affective and cognitive investment, which jointly propelled behavioral engagement (Boekaerts, 2016). Another factor may be participants' awareness of higher-order issues. Some participants prioritized content and organization of writing products, which possibly promoted their serious treatment of higher-order issues mentioned in accurate and inaccurate feedback. Moreover, the second argumentative essay in our study accounted for 7.5% of the course mark, thus probably triggering most participants' behavioral devotion to various feedback. Undeniably, our study also reveals that behavioral engagement may conceal learner's deficient cognitive activity, which indicates the possible inconsistency of the two dimensions (Fan & Xu, 2020).

7. Conclusion

Given the significance of student engagement in effective use of genAI feedback for improving EFL writing, student engagement with Kimi feedback was gauged

from the tripartite lens with a class of college-level EFL students as participants. The research results shed light on the characteristics of Kimi feedback pattern in its multiple levels and high accuracy. More importantly, the results provide a comprehensive picture of EFL students' positive engagement with Kimi feedback affectively, cognitively, and behaviorally. Meanwhile, the potential inconsistency exists between cognitive and behavioral dimensions. The research proves the multi-faceted nature of student engagement with genAI feedback in L2 writing activity. It also highlights a multi-faceted lens for teachers to comprehend students' response to genAI feedback and to offer tailored scaffolds for students' more effective utilization of such affordance.

Several limitations in our study should be acknowledged. Our participants have upper-intermediate English proficiency, whose engagement with genAI feedback possibly differs from that of learners at other proficiency levels. Therefore, this limits the generalizability of our findings.

to other situations with a wider learner population. Further research can be conducted with diverse samples at intermediate or low English proficiency level to track their involvement. Besides, our study chooses Kimi chatbot for its easy availability to our participants. Yet feedback produced by different genAI tools may present discrepancy in accuracy, clarity, etc., which perhaps leads to variation in student engagement. Consequently, researchers are suggested to probe other genAI tools to expand the knowledge about the characteristics of student engagement with this affordance.

Author Contributions

Xiaohua Yang: designing and conducting the research, and writing the paper. Xiumei Liu: designing the research, checking the accuracy of the interview transcripts, checking the translation of the transcript and reflection papers.

Funding

This research was supported by funding from the 12th China Foreign Language Education Fund of National Research Center of Foreign Language Education, Beijing Foreign Studies University (Award Number: ZGWYJYJJ12A145); the Teaching Reform Project of China University of Mining and Technology-Beijing (Award Number: J240806).

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Abendschein, B., Lin, X., Edwards, C., Edwards, A., & Rijhwani, V. (2024). Credibility and Altered Communication Styles of AI Graders in the Classroom. *Journal of Computer*

- Assisted Learning*, 40, 1766-1776. <https://doi.org/10.1111/jcal.12979>
- Annamalai, N., Rashid, R. A., Hashmi, U. M., Mohamed, M., Alqaryouti, M. H., & Sadeq, A. D. (2023). Using Chatbots for English Language Learning in Higher Education. *Computers and Education: Artificial Intelligence*, 5, Article ID: 100153. <https://doi.org/10.1016/j.caeai.2023.100153>
- Ashwell, T. (2000). Patterns of Teacher Response to Student Writing in a Multiple-Draft Composition Classroom: Is Content Feedback Followed by Form Feedback the Best Method? *Journal of Second Language Writing*, 9, 227-257. [https://doi.org/10.1016/s1060-3743\(00\)00027-8](https://doi.org/10.1016/s1060-3743(00)00027-8)
- Barrios-Beltran, D. (2025). Exploring the Efficacy of ChatGPT-4 Feedback in Second Language Spanish Writing. *System*, 133, Article ID: 103771. <https://doi.org/10.1016/j.system.2025.103771>
- Barrot, J. S. (2023). Using ChatGPT for Second Language Writing: Pitfalls and Potentials. *Assessing Writing*, 57, Article ID: 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Boekaerts, M. (2016). Engagement as an Inherent Aspect of the Learning Process. *Learning and Instruction*, 43, 76-83. <https://doi.org/10.1016/j.learninstruc.2016.02.001>
- Chen, M., & Lv, M. (2024). College English Writing Instruction in a ChatGPT Context. *Contemporary Foreign Language Studies*, 24, 161-168.
- Chen, W. (2019). An Exploratory Study on the Role of L2 Collaborative Writing on Learners' Subsequent Individually Composed Texts. *The Asia-Pacific Education Researcher*, 28, 563-573. <https://doi.org/10.1007/s40299-019-00455-3>
- Cummings, R. E., Monroe, S. M., & Watkins, M. (2024). Generative AI in First-Year Writing: An Early Analysis of Affordances, Limitations, and a Framework for the Future. *Computers and Composition*, 71, Article ID: 102827. <https://doi.org/10.1016/j.compcom.2024.102827>
- Ellis, R. (2010). EPILOGUE: A Framework for Investigating Oral and Written Corrective Feedback. *Studies in Second Language Acquisition*, 32, 335-349. <https://doi.org/10.1017/s0272263109990544>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-Generated Feedback on Writing: Insights into Efficacy and ENL Student Preference. *International Journal of Educational Technology in Higher Education*, 20, Article No. 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Fan, Y., & Xu, J. (2020). Exploring Student Engagement with Peer Feedback on L2 Writing. *Journal of Second Language Writing*, 50, Article ID: 100775. <https://doi.org/10.1016/j.jslw.2020.100775>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74, 59-109. <https://doi.org/10.3102/00346543074001059>
- Guo, K., & Wang, D. (2023). To Resist It or to Embrace It? Examining ChatGPT's Potential to Support Teacher Feedback in EFL Writing. *Education and Information Technologies*, 29, 8435-8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-Supported Approach to Peer Feedback on University EFL Students' Feedback Quality and Writing Ability. *The Internet and Higher Education*, 63, Article ID: 100962. <https://doi.org/10.1016/j.iheduc.2024.100962>
- Han, Y., & Hyland, F. (2015). Exploring Learner Engagement with Written Corrective Feedback in a Chinese Tertiary EFL Classroom. *Journal of Second Language Writing*, 30, 31-44. <https://doi.org/10.1016/j.jslw.2015.08.002>

- Hyland, F. (1998). The Impact of Teacher Written Feedback on Individual Writers. *Journal of Second Language Writing*, 7, 255-286.
[https://doi.org/10.1016/s1060-3743\(98\)90017-0](https://doi.org/10.1016/s1060-3743(98)90017-0)
- Koltovskaia, S. (2020). Student Engagement with Automated Written Corrective Feedback (AWCF) Provided by Grammarly: A Multiple Case Study. *Assessing Writing*, 44, Article ID: 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate Students' Use of ChatGPT for Academic Text Revision: Behavioral, Cognitive, and Affective Engagement. *Journal of Second Language Writing*, 65, Article ID: 101130.
<https://doi.org/10.1016/j.jslw.2024.101130>
- Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (2nd ed.). Sage.
- Langan, J. (2014). *College Writing Skills with Readings* (9th ed.). Foreign Language and Teaching and Research Press.
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating Students' Cognitive Processes in Generative AI-Assisted Digital Multimodal Composing and Traditional Writing. *Computers & Education*, 211, Article ID: 104977.
<https://doi.org/10.1016/j.compedu.2023.104977>
- Long, M. H. (1996). The Role of the Linguistic Environment in Second Language Acquisition. In W. Ritchie, & T. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). Elsevier. <https://doi.org/10.1016/b978-012589042-7/50015-3>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A. et al. (2024). Using LLMs to Bring Evidence-Based Feedback into the Classroom: AI-Generated Feedback Increases Secondary Students' Text Revision, Motivation, and Positive Emotions. *Computers and Education: Artificial Intelligence*, 6, Article ID: 100199.
<https://doi.org/10.1016/j.caeai.2023.100199>
- Miles, M. B., Huberman, M. A., & Saldaña, J. (2020). *Qualitative Data Analysis: A Methods Sourcebook* (4th ed.). Sage.
- Mizumoto, A., & Eguchi, M. (2023). Exploring the Potential of Using an AI Language Model for Automated Essay Scoring. *Research Methods in Applied Linguistics*, 2, Article ID: 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Muñoz Muñoz, B. C., Nassaji, H., & Bello Carrillo, F. I. (2025). ChatGPT-Generated versus Human Direct Corrective Feedback on L2 Writing. *System*, 134, Article ID: 103805.
<https://doi.org/10.1016/j.system.2025.103805>
- Ou, A. W., Stöhr, C., & Malmström, H. (2024). Academic Communication with AI-Powered Language Tools in Higher Education: From a Post-Humanist Perspective. *System*, 121, Article ID: 103225. <https://doi.org/10.1016/j.system.2024.103225>
- Prapahan, P. W., & Prapahan, K. (2023). AI Technologies in the ESL/EFL Writing Classroom: The Villain or the Champion? *Journal of Second Language Writing*, 62, Article ID: 101072. <https://doi.org/10.1016/j.jslw.2023.101072>
- Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using Artificial Intelligence to Foster Students' Writing Feedback Literacy, Engagement, and Outcome: A Case of Wordtune Application. *Interactive Learning Environments*, 32, 5020-5040.
<https://doi.org/10.1080/10494820.2023.2208170>
- Ruwe, T., & Mayweg-Paus, E. (2023). "Your Argumentation Is Good", Says the AI vs Humans—The Role of Feedback Providers and Personalised Language for Feedback Effectiveness. *Computers and Education: Artificial Intelligence*, 5, Article ID: 100189.
<https://doi.org/10.1016/j.caeai.2023.100189>

- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in Argumentative Writing Classrooms. *Assessing Writing*, 57, Article ID: 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Sun, Q., Pan, H., & Zhan, J. (2023). Untangling the Relationship between English as a Foreign Language Learners' Metacognitive Experiences and Writing Proficiency: A Mixed-Methods Approach. *System*, 117, Article ID: 103100. <https://doi.org/10.1016/j.system.2023.103100>
- Tian, L., & Zhou, Y. (2020). Learner Engagement with Automated Feedback, Peer Feedback and Teacher Feedback in an Online EFL Writing Context. *System*, 91, Article ID: 102247. <https://doi.org/10.1016/j.system.2020.102247>
- Tseng, W., & Warschauer, M. (2023). AI-Writing Tools in Education: If You Can't Beat Them, Join Them. *Journal of China Computer-Assisted Language Learning*, 3, 258-262. <https://doi.org/10.1515/jccall-2023-0008>
- Tsui, A. B. M., & Ng, M. (2000). Do Secondary L2 Writers Benefit from Peer Comments? *Journal of Second Language Writing*, 9, 147-170. [https://doi.org/10.1016/s1060-3743\(00\)00022-9](https://doi.org/10.1016/s1060-3743(00)00022-9)
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
- Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing Argumentative Writing with Automated Feedback and Social Comparison Nudging. *Computers & Education*, 191, Article ID: 104644. <https://doi.org/10.1016/j.compedu.2022.104644>
- Wang, X., & Wang, S. (2024). Exploring Chinese EFL Learners' Engagement with Large Language Models: A Self-Determination Theory Perspective. *Learning and Motivation*, 87, Article ID: 102014. <https://doi.org/10.1016/j.lmot.2024.102014>
- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The Affordances and Contradictions of AI-Generated Text for Writers of English as a Second or Foreign Language. *Journal of Second Language Writing*, 62, Article ID: 101071.
- Wegerif, R. (2004). The Role of Educational Software as a Support for Teaching and Learning Conversations. *Computers & Education*, 43, 179-191. <https://doi.org/10.1016/j.compedu.2003.12.012>
- Wei, S., & Li, L. (2023). Artificial Intelligence-Assisted Second Language Writing Feedback: A Case Study of ChatGPT. *Foreign Languages in China*, 20, 33-40.
- Wen, Q., & Liang, M. (2024). Human-AI Interactive Negotiation Competence: ChatGPT and Foreign Language Education. *Foreign Language Teaching and Research*, 56, 286-296.
- Wiboolyasarini, W., Wiboolyasarini, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing Collaborative Writing and AI Feedback: An Investigation into Enhancing L2 Writing Proficiency in Wiki-Based Environments. *Computers and Education: Artificial Intelligence*, 6, Article ID: 100228. <https://doi.org/10.1016/j.caeai.2024.100228>
- Yang, X., & Xu, H. (2022). Effects of Mapping as a Prewriting Strategy Training for Peer Written Feedback. *Foreign Languages and Their Teaching*, No. 1, 84-94. (In Chinese)
- Yu, S., Zhang, Y., Zheng, Y., Yuan, K., & Zhang, L. (2019). Understanding Student Engagement with Peer Feedback on Master's Theses: A Macau Study. *Assessment & Evaluation in Higher Education*, 44, 50-65. <https://doi.org/10.1080/02602938.2018.1467879>
- Zhang, Z. & Hyland, K. (2018). Student Engagement with Teacher and Automated Feedback on L2 Writing. *Assessing Writing*, 36, 90-102. <https://doi.org/10.1016/j.asw.2018.02.004>

- Zhang, Z. (2020). Engaging with Automated Writing Evaluation (AWE) Feedback on L2 Writing: Student Perceptions and Revisions. *Assessing Writing*, 43, Article ID: 100439. <https://doi.org/10.1016/j.asw.2019.100439>
- Zhang, Z., & Hyland, K. (2025). The Role of Digital Literacy in Student Engagement with Automated Writing Evaluation (AWE) Feedback on Second Language Writing. *Computer Assisted Language Learning*, 38, 1060-1085. <https://doi.org/10.1080/09588221.2023.2256815>
- Zheng, Y., & Yu, S. (2018). Student Engagement with Teacher Written Corrective Feedback in EFL Writing: A Case Study of Chinese Lower-Proficiency Students. *Assessing Writing*, 37, 13-24. <https://doi.org/10.1016/j.asw.2018.03.001>

Appendix

Appendix A. Scoring rubric for argumentative essay based on Guo et al. (2024), Chen (2019), and Langan (2014: p. 151)

Content	Argument in its clarity	2 points
	Argument development with related, convincing, and varied details	2 points
	Effective addressing of the counterargument	2 points
Organization	A clear and engaging introduction	2 points
	Logical and integrated body paragraphs: topic sentences leading body paragraphs, logical consistence among all the supporting details and sentences in an essay	2 points
	A powerful conclusion including restatement of the thesis and suggestion, prediction, etc.	2 points
	The use of transitions between ideas and paragraphs	2 points
Form	Lexical (complex range, accurate word and idiom choice, mastery of word forms, appropriate register)	2 points
	Grammatical (complex structures, agreement, number, tense, word order, articles, pronouns, prepositions)	2 points
	Mechanical (spelling, punctuation, capitalization, paragraph indentation, miscellaneous issue)	2 points
Total		20 points

Appendix B. Prompt handout for student-AI communication

- 1) Is the argument of my essay clearly stated? How can I clarify it?
- 2) Do I provide related, convincing, and varied evidence to support the argument?
- 3) Is there any point off-topic in my essay? What areas need more details?
- 4) Where can I improve the counterargument?
- 5) Do I have a clear introduction?
- 6) Does my conclusion effectively summarize my main points and restate the thesis?
- 7) Do my topic sentences clearly convey the main idea of each paragraph? How can I clarify them?
- 8) Do the sentences flow logically from one to the next? Are there any abrupt transitions?
- 9) Do my paragraphs flow logically from one to the next? Are there any abrupt transitions?
- 10) Do I use a variety of words?
- 11) Are there words that could be replaced with more accurate alternatives?
- 12) Do I use the words in their correct form?
- 13) Are there words that could be replaced to maintain a formal style?
- 14) Do I use a variety of sentence structures? Are there any repetitive patterns I should avoid?

15) Are there any grammatical errors that I should correct? Explain them, and provide correction.

16) Are there any mechanical errors that I should correct? Explain them, and provide correction.

Appendix C. Prompt questions for reflection paper

Please reflect on what you have experienced in your using Kimi feedback.

- 1) What did you do when you were provided various Kimi feedback?
- 2) Could you understand Kimi feedback to your essay?
- 3) What methods did you use to understand Kimi feedback and to decide whether to accept it?
- 4) What did you do after your revision?

Chinese version

请回顾你使用 Kimi 反馈的经历。

- 5) 面对 Kimi 给出的多个反馈，你是怎样应对的？
- 6) 你能理解 Kimi 对你文章的反馈吗？
- 7) 你是采用什么方法理解 Kimi 反馈和决定是否采纳 Kimi 反馈的？
- 8) 修改之后你会做什么？

Appendix D. Questionnaire about participants' affective engagement with Kimi feedback

- 1) I am interested in the writing feedback functionality of Kimi.
- 2) I was willing to discuss my essay with Kimi.
- 3) I felt happy after reading Kimi feedback to my essay.
- 4) I felt embarrassed after reading Kimi feedback to my essay.
- 5) Kimi feedback was useful in improving the content of my English essay.
- 6) Kimi feedback was useful in improving the organization of my English essay.
- 7) Kimi feedback was useful in improving the form of my English essay.

Chinese version

- 1) 我对 Kimi 的作文反馈功能感兴趣。
- 2) 我愿意和 Kimi 讨论我的作文。
- 3) 看到 Kimi 对我作文的反馈我觉得愉快。
- 4) 看到 Kimi 对我作文的反馈我觉得反感。
- 5) Kimi 反馈对提高我的英语作文的内容有用。
- 6) Kimi 反馈对提高我的英语作文的结构有用。
- 7) Kimi 反馈对提高我的英语作文的形式有用。

Appendix E. Guiding questions for semi-structured interview

- 1) What do you think of the feedback functionality of Kimi?
- 2) How did you feel when you receive Kimi feedback?
- 3) Can you understand Kimi feedback?
- 4) What is the influence of Kimi feedback on your English writing?
- 5) What did you do after receiving Kimi feedback?
- 6) Are you willing to receive genAI feedback in future writing activity?
- 7) Do you have any suggestions for using genAI feedback in future writing activity?

Chinese Version

- 1) 你怎样看待 Kimi 的反馈功能?
- 2) 得到 Kimi 反馈后你的感受是什么?
- 3) 你能理解 Kimi 的反馈吗?
- 4) Kimi 反馈对你的英语写作有什么影响?
- 5) 收到 Kimi 反馈后你是怎么做的?
- 6) 你愿意在未来的写作活动中得到生成式人工智能反馈吗?
- 7) 你对在写作活动中使用生成式人工智能反馈有什么建议吗?