Research Article

Unpacking the Rejection of L2 Students Toward ChatGPT-Generated Feedback: An Explanatory Research

Ziqi Chen (陈子淇) iD
Macao Polytechnic University

Wei Wei (韦卫)
Macao Polytechnic University

Xinhua Zhu (祝新华)
The Hong Kong Polytechnic University

Yuan Yao (姚远) iD
The Hong Kong Polytechnic University

Abstract

Purpose: Drawing on the Unified Theory of Acceptance and Use of Technology (UTAUT), this sequential mixed-method study explored L2 learners' challenges in seeking corrective feedback in an AI-assisted educational environment.

Design/Approach/Methods: The data was collected from 45 university students in a Computer Science program who were encouraged to seek corrective feedback from ChatGPT for their argumentative writing report. Self-reflection data were qualitatively and quantitatively analyzed.

Findings: Findings suggested that 45.9% of the AI-generated feedback was rejected by students, with a higher rejection rate observed for content-focused feedback (58.7%) compared to form-focused feedback (41.3%). AI-generated content-focused feedback was rejected due to misalignments between expected and received feedback, anticipated high workload to respond, mismatch

Corresponding author:
Wei Wei, Faculty of Applied Sciences, Macao Polytechnic University, Room P325, Pearl Jubilee Building, Taipa, Macao.
Email: weiweitesting@hotmail.com

between feedback from AI and external references, and impeding conditions for students to engage with the feedback. Meanwhile, the rejection of form-focused feedback was largely associated with the high level of expected effort.

Originality/Value: This study is one of the few existing studies on the L2 learners' challenges in seeking corrective feedback from generative AI systems. Notably, the study differentiates between form-focused and content-focused feedback and identifies the distinct challenges and learning opportunities that impact students' rejection of AI feedback.

Introduction

Technology integration with classroom instructions and informal learning outside of the classroom has been investigated for decades (Hinostroza et al., 2015; Lai, 2015; Morales & Roig, 2002). The technology of Open AI and chatbots present extraordinary potential in promoting better language learning and teaching practice, such as timely feedback, ease of use, personalized support, interlocutors, simulations of the learning environment, the transmission of information, helpline, and individualized recommendation on learning materials (Huang et al., 2022; Lu et al., 2024).

However, its limitations and challenges to support teaching and learning cannot be overlooked. Reviewing the challenges in previous studies, this study aims to conduct an in-depth analysis of the feedback exchange process between students and ChatGPT-3.5. Through the research identifying the challenges that arise during this engagement and their potential impact on student behaviors, we aim to gain valuable insights into how these obstacles, as perceived by students, impede the effective acceptance and utilization of feedback.

Literature review

Numerous research studies have examined the challenges of integrating distinct technology in educational settings (Bećirović, 2023). However, studies that explore the challenges of integrating Generative AI in academic writing are still limited. With the new launch of ChatGPT, a generative AI-assisted chatbot, we can anticipate potential limitations by analyzing previous research on chatbot-assisted language learning and teaching.

Technological limitations

Technology itself is reported as the challenge and barrier to technology integration with classroom instruction and self-directed learning activities. It includes (1) incapability to offer answers and responses in a reliable and expected way (Kuhail et al., 2023; Winkler et al., 2020), (2) unavailability of emotional support to the users (Escalante et al., 2023; Gallacher et al., 2018; Goda et al., 2014), and (3) limitation in engaging and maintaining conversations (Yang & Zapata-Rivera, 2010).

In relation to the first point, chatbots may not be able to accurately respond to users' inquiries due to the limitations or insufficiency of training datasets in the early stage of the development of the AI systems. For example, Winkler et al. (2020) designed a conversational bot named Sara aimed to solve the challenge of creating meaningful interactions in online video lectures. They evaluated the bot by collecting self-reflection reports from 182 participants who were undergraduate and graduate students. The findings indicated that bots with limited datasets posed challenges to learning, including the inability to provide correct answers and respond to questions. These limitations led to learners' frustrations and thus possibly influenced their learning process.

Furthermore, chatbots lack human-like emotional support, which can be linked to unnatural and ambiguous voices and the use of non-academic language. According to Goda et al.'s (2014) study, students have reported feeling that computer-generated voices are unnatural compared to human voices, ultimately limiting their engagement with the immersive learning environment. Moreover, Gallacher et al. (2018) noted that the absence of emotional and visible cues during interactions with chatbots undermined students' positive state of language learning. While Escalante et al. (2023) collected 48 undergraduate students' perceptions of feedback provided by teachers and feedback given by generative AI, the result revealed that half preferred feedback provided by teachers rather than AI-generated feedback. The primary reason cited for this preference was the positive emotional impact of in-person communication.

Lastly, the limitation of chatbots in engaging in extended conversations is another drawback. Research by Yang and Zapata-Rivera (2010) revealed that students' ideas diverge in unforeseen directions as communication progresses, and chatbot systems fail to recognize and address these constantly shifting topics.

The challenges and barriers to teachers and learners

Regarding chatbot limitations in education, a subset of research has primarily focused on teachers and learners. Four areas were identified from the perspective of teachers: (1) teachers' negative beliefs and attitudes toward the limited use of technology in teaching (Howard, 2013; Shin et al., 2014), (2) teachers' discomfort or lack of confidence in working with technology, which can lead to the rejection of new technology (Leem & Sung, 2019), (3) teachers' low level of

computer literacy can make it difficult for them to operate chatbot technology (Su et al., 2023), and (4) limited access to resources in schools can also pose a challenge that can hinder the implementation of AI tools (Reiser, 2002; Shin et al., 2014).

More importantly, from the learners' perspective, several factors may be relevant to the limitations of implementing new technologies in teaching and learning: (1) students' low perceived usefulness can lead to negative behaviors during their interactions with the technology (Chen et al., 2020; Kaur et al., 2021; Marchante, 2022), (2) students' limited competence and computer literacy may hinder their effective use of chatbot technology (Celik, 2023; Hatlevik et al., 2018; Fryer et al., 2020), and (3) the disconnect between students' external requirements and their interactions with technology has resulted in minimal use of chatbots in their learning process (Ericsson et al., 2023; Guo et al., 2023; Hew et al., 2023).

To be more specific, students' attitudes and beliefs toward the use of technology in supporting their learning process refer to their technological acceptance. This encompasses their willingness to experiment with new technology, perceived usefulness, and overall attitudes toward technology as an educational tool. For instance, Chen et al. (2020) found that students' beliefs about chatbot usefulness in Chinese language learning significantly impacted their acceptance and utilization of the technology. Moreover, an insightful study by Hatlevik et al. (2018) examined the connection between technology and students' self-efficacy in nearly 60,000 individuals from 15 countries. The results revealed that familiarity with technology significantly influenced students' information and communication technology self-efficacy. Fryer et al.'s (2020) study revealed that learners are unlikely to develop an interest in oral tasks during language classes without sufficient training in feedback-seeking strategies and computer literacy. As a result, human-supported oral training sessions achieved better outcomes, with higher levels of motivation and engagement. Furthermore, the successful implementation of chatbots in educational environments depends on the alignment between students' learning interest in AI and assignment requirements established by educators (Hew et al., 2023).

Framework and research questions

Currently, no existing research utilizes a theoretical framework to systematically analyze the challenges associated with chatbot-assisted writing and corrective feedback practices. This study aims to fill that gap by utilizing the Unified Theory of Acceptance and Use of Technology (UTAUT), first introduced by Venkatesh et al. (2003). Upon examining the breadth of research on Generative AI within educational settings, we found that the UTAUT model elucidates multiple barriers to the optimal and efficient deployment of these AI systems. The model identifies four principal areas impacting effectiveness and integration, which have already been widely discussed separately in the field of AI-assisted learning and teaching. Firstly, there is "Performance Expectancy" related

to the extent to which AI can visualize and be assisted in achieving learning outcomes. For instance, previous studies have shown that AI is a reliable tool for scoring and peer review (Jiang et al., 2023; Kaebnick et al., 2023). Secondly, "Effort Expectancy" considers the ease with which these tools can be used to facilitate the learning process, and sometimes even in an unethical manner, such as plagiarism and academic integrity issues (Cotton et al., 2023). Thirdly, "Social Influence" is tied to the endorsement and contempt from other stakeholders in the educational community. McDonald et al. (2024) reviewed the policies of 116 US higher education institutions on the use of Generative AI, and most of the policy concerns were issues of ethical practice of AI, with most references to its application remaining ambiguous. Lastly, "Facilitating Conditions" refers to the infrastructural and resource-related aspects that enable AI utilization. As indicated in prior research (Reiser, 2002; Shin et al., 2014), constraints related to accessing necessary resources in academic settings can greatly hinder the successful implementation and effective use of AI technologies. In our study, the four core factors are "Mismatch with learners' expectations," "Anticipated high workload to respond," "Mismatch with external references," and "Impending conditions." We have provided definitions for each of these core factors in Table 2.

These potential challenges can be suggested by observing students' revision strategies to deal with AI-generated feedback. As for the framework for coding students' revision strategies, two categories are identified from previous corrective feedback studies based on the rate of uptake or acceptance of received feedback (Cheong et al., 2023): (1) accept and revise accordingly and (2) show a certain level of rejection, such as ignorance, disagreement, and confusion by arguing with the feedback providers, especially for non-error corrective feedback (Mohebbi, 2021; Yang et al., 2023; Zhang, 2020). Moreover, we used two established frameworks, called "levels of revisions" and "taxonomy of revision strategies," to label the dependent variable. These frameworks were informed by prior research on corrective feedback and automatic written feedback. Fan and Xu's (2020) research shows two types of feedback: form-focused and content-focused. Form-focused feedback, typically given by teachers or peers, targets errors in language usage, such as sentence structures, vocabulary, and verb conjugations. Interestingly, Fan and Xu (2020) found that content-focused feedback was primarily delivered through negotiation. This means that feedback providers prefer to engage in oral discussions and debates with receivers on topics related to evidence and examples.

Without any doubt, partnerships with Generative AI technology in language teaching and learning will be one of the trends. The use of AI technologies in language teaching and learning has gained significant attention in recent years (Hinostroza et al., 2015; Lai, 2015; Morales & Roig, 2002). Although the digital affordance of these AI-driven learning systems has been widely investigated (Fitria, 2021; Ghufron & Rosyida, 2018), the challenges of using Generative AI technologies have not been adequately examined. Open AI released a powerful generative AI technology

called ChatGPT in late November 2022. To date, most of the research related to ChatGPT has mainly concentrated on academic writing (Rozencwajg & Kantor, 2023; Salvagno et al., 2023; Zimmerman, 2023). However, an in-depth exploration of ChatGPT's distinctive influence on enhancing student engagement and their responses to formative corrective feedback is notably limited.

In that case, this research aims to incorporate ChatGPT technology into the language learning feedback process. Guided by the UTAUT, the study aims to identify the key factors that lead to negative responses to ChatGPT feedback. Consequently, the study is guided by two primary questions:

1. What are the challenges faced by L2 learners when engaging with ChatGPT-generated feedback?
2. What are the underlying factors that contribute to L2 learners' rejection of form- and content-focused feedback generated by ChatGPT?

Methods

In this study, we used an exploratory sequential mixed methods design to investigate the use and reception of AI-generated feedback in academic writing among university students. The study was conducted with 45 students enrolled in a computer science program who were encouraged to seek corrective feedback from ChatGPT on their argumentative writing reports. In an initial qualitative phase, self-reflective data was collected from the students, providing a rich, contextual understanding of their experiences and reactions to the AI-generated feedback. We subsequently conducted a quantitative analysis of the data to extrapolate the initial findings to the wider student population.

Participants

Undergraduate students from a Computer Science program at a public university in Macao, aged between 20 and 21 ($SD = 1.72$), were invited to participate in an academic writing course. The sample consisted of 52 male and seven female students who had all completed the first-semester courses and understood computer operational systems, AI technologies, and English academic writing well. The language of instruction in the classroom was English, so the participants felt comfortable communicating with the AI system in English. Among the sample, 55 were native Cantonese or Mandarin speakers, and the remaining four were Portuguese, English, and Filipino native speakers. Most participants were at the B1 or B2 level at the CEFR band scale (Common European Framework of Reference for Languages), based on their scores in the English language test of the university entrance examination. At the time of data collection, most participants had

already used ChatGPT, the chatbot, for academic and entertainment purposes since its official launch at the end of November 2022. These students were first contacted to participate in this study via Canvas, a centralized online learning system. Forty-five students expressed interest in the study, and written consent forms were obtained from each of them. Data protection strategies included removing all personal identifiers and replacing them with ID numbers.

Data collection

After that, these students were briefed about the procedure of data collection. First, they were invited to draft a 250-word essay individually without the assistance of ChatGPT in a classroom condition. The writing task is "To what extent do you agree or disagree that AI technologies may replace human working forces?". This task was selected for several reasons: (1) this topic has been discussed in one of their compulsory subjects in their first semester, Ethics and Professional Issues, and (2) they were familiar with relevant ideas, scientific reports, and linguistics resources. Ethics approval was obtained by the organization of the corresponding author. The data collection for this study is an integral part of our routine work. No additional academic work is placed on participants specifically for this study.

Secondly, they were trained to use ChatGPT to offer corrective feedback on their draft and revise their essay in 2-week advance. The training includes (1) the academic integrity issues in the use of ChatGPT to complete academic assignments, (2) the university guidelines on the use of ChatGPT for academic purposes, and (3) the guidelines and examples of writing reflective journals to report their feedback experiences with ChatGPT.

The participants in the study were asked to interact with ChatGPT in two ways: (1) by providing corrective feedback on their essay, which they attached, and (2) by making suggestions to improve their essay based on the attached marking rubrics. These rubrics were designed to evaluate both the form-focused aspects of writing (grammatical accuracy and lexical sophistication) as well as the content-focused aspects (quality of argument structures and use of source evidence) (Wei et al., 2022). Participants were encouraged to ask ChatGPT follow-up questions about its feedback, including asking for clarifications, providing more examples, or even disputing the feedback if they disagreed. Finally, they were required to submit their responses to ChatGPT's feedback item by item in either their first language or English. The collected data included the participants' first draft of their essays, the corrective feedback and suggestions from ChatGPT, and the learners' responses and revisions, as shown in Table 1.

Data coding

To answer the first research question, the UTAUT model was adapted and developed to identify and code possible barriers to students' negative responses. The first author and corresponding author

Table 1. Example of L2 writers' response on ChatGPT generated feedback.

Feedback from ChatGPT	Your response
1. Use proper grammar and punctuation—here are several grammatical errors in the essay, including improper use of commas and run-on sentences. Make sure to proofread your work before submitting it.	I agreed that, and I provide made revision accordingly. I change "get in touch with people" to "easily connect with people", "had convenient our daily lives" to "has made our daily lives more convenient", "book for position" to "make reservations".
2. Use appropriate language—Choose your words carefully and use appropriate language for your audience and purpose. Avoid using overly technical terms or jargon that your readers may not understand. Also, make sure to vary your sentence structure to keep your writing interesting.	I don't agree that, think my words are all very easy. I don't use any jargon in my essay. The evidence I chosen were closed to our lives.
3. Clarify and elaborate on your points—Some of your arguments could be more specific and detailed. For example, you could provide examples of jobs that have been replaced by machines or describe how addiction to virtual technology affects young people's lives.	I agree with that, but I can't do so. Because that will over the word count.

jointly coded the responses of the first 15 participants and developed the framework collaboratively. After resolving all disagreements, they coded the rest of the responses from the remaining 30 participants with an agreement rate of 87%. The specific explanation of coding factors is presented in Table 2.

To answer the second research question, an Ordinal Logistic Regression model (OLR) was developed to estimate the prediction power of the identified barriers to learners' rejection of the AI-generated feedback. To label the dependent variable, the coding process was guided by two established frameworks: levels of revisions and the taxonomy of revision strategies, both informed by prior research on corrective feedback and automatic written feedback. This study considered four critical, independent variables in this model, including "Mismatch with learners' expectations," "Anticipated high workload to respond," "Mismatch with external references," and "Impeding conditions" to predict users' intentions to engage with ChatGPT-generated feedback. The OLR model can create an analysis model that treats the dependent variable as an ordered categorical variable based on the cumulative odds principle, similar to a conventional multiple regression approach (Hosmer et al., 2013).

Table 2. Coding framework on the students' negative responses to ChatGPT-generated feedback.

Categories	Explanations
Mismatch with learners' expectations	ChatGPT provided unhelpful feedback, such as (1) misinterpreting authors' intentions and ideas, (2) lack of clarity and examples, and (3) extraneous feedback, which made students believe using this technology would not help them attain work/study performance gains.
Anticipated high workload to respond	ChatGPT provided inconvenient feedback, such as (1) information overload in feedback and (2) a general description rather than highlighting specific mistakes, which decreased students' degree of ease associated with using the particular technology in their work.
Mismatch with external references	ChatGPT gave feedback that did not align with the expectations or preferences of other important people (e.g., peers and teachers), making students refuse to take the AI feedback.
Impeding conditions	ChatGPT did not provide personalized suggestions based on proficiency levels and lacked emotional support, causing discomfort and unexpected behavior in students.

Result

Qualitative findings

To answer the first research question, the research conducted a thorough analysis of the feedback provided by L2 learners to understand their rejection of AI-generated feedback better. We carefully examined the transcript, identified relevant themes that could serve as suitable study questions, and answered the first research question. By analyzing 167 L2 students' rejections of generative AI feedback, we found that students rejected the feedback more often because of "Mismatch with learners' expectations" (37.1%) and "Anticipated high workload to respond" (34.1%) than "Mismatch with external references" (15.0%) and "Impeding conditions" (13.8%). Specifically, the findings have been summarized in Table 3, which provides an overview of the barriers encountered by L2 learners when attempting to act upon ChatGPT-generated feedback guided by the UTAUT model.

The "Mismatch with learners' expectations" aspect of feedback refers to the feedback provided by ChatGPT that led students to think that using this technology would not help them improve their work or study performance. First, ChatGPT may not fully understand the nuances and subtleties of the learners' writing. As a result, it may misinterpret the authors' intentions and make revisions that deviate from the original meaning. For example:

> I think ChatGPT's comment is wrong because I just want to speak from a neutral point of view to show some good and bad things coming out. Therefore, I will not make any changes on this point.

Table 3. The reasons behind students' rejection of AI feedback guided by the UTAUT model.

Factors of the UTAUT model	Sub-category
Mismatch with learners' expectations	misinterpreting authors' intentions and ideas (31)
	lack of clarity and examples (21)
	extraneous feedback (10)
Anticipated high workload to respond	a general description rather than highlighting specific mistakes (33)
	information overload (24)
Mismatch with external references	not aligned with other teachers' and peers' expectations or preferences (25)
Impeding conditions	lack of personalized suggestions based on proficiency levels (18)
	lack of emotional support (5)

Nevertheless, the learners pointed out that the suggestions lacked clarity and specific examples, which in turn made it challenging for them to identify and address their particular areas of improvement, as one student stated:

> I do not understand what "changed the formatting" means; therefore, I cannot make revisions.

In ChatGPT, extraneous feedback refers to the frustrating experience of L2 learners when they receive irrelevant corrections from AI models meant to assist them. Extraneous feedback frustrates L2 learners who use AI language models for assistance, wasting time and confusion. For example:

> I think ChatGPT's comment is wrong because my first sentence of the paragraph is, "On the other hand, technology also has some disadvantages." And ChatGPT gave me the same sentence.

The "Anticipated high workload to respond" challenges in AI feedback decreased students' degree of ease in using the particular technology in their work. For instance, overload-specific suggestions given by ChatGPT are often disregarded and rejected by the learners. The abundance of tips can overwhelm learners and cause them to overlook valuable insights or fail to allocate their attention appropriately. For example:

> Its summary is too long to be suitable for its summary. It will be better if it gives me a short one.

> Yes, I agree with this feedback. I use some examples to support my arguments. For example, when a student has questions, WeChat software can help students connect; if the teacher has something important to say, it can also help the teacher send the information to students faster. (However, the ChatGPT suggested to put more statistics, quotes, or research studies in writing.)

Furthermore, ChatGPT provided a generic explanation of an error, resulting in L2 learners failing to make the necessary revisions and seeking clarification on the specific occurrences of these errors. This challenge may hinder the learners' ability to recognize and rectify their linguistic shortcomings by not highlighting the exact locations of errors. For example:

> ChatGPT did not give me the exact sentences, like which sentence is unclear; it just told me to revise all the paragraphs, so I will not make any changes on this point.

The "Mismatch with external references" challenge in AI feedback refers to ChatGPT providing feedback that does not align with the expectations or preferences of essential people, such as peers and teachers, leading students to reject the AI feedback from ChatGPT despite its accuracy and helpfulness. Additionally, students may prioritize the feedback of individuals they trust over AI feedback, as one student mentioned:

> I think ChatGPT is wrong and right. More examples can support a convincing essay. Nevertheless, as I write above, I should follow the requirements of the $3 + 7 + 7 + 3$. So, for all the advantages and disadvantages, I can only add one example to each of them.

Our research identified challenges concerning "Impeding conditions" related to students' reluctance to use AI feedback in their essays. Specifically, ChatGPT could not provide personalized suggestions based on proficiency levels and lacked emotional support, leading to discomfort and negative behavior in some students. This challenge might have made the learning process tedious and demotivating for students, as some students stated:

> I agree with that, but I cannot do much to avoid it because of my weak English skills.

> I also want to express myself in simpler language, but my vocabulary is small, and I can only express myself as much as possible.

Quantitative findings

In order to answer the second research question, descriptive statistical analysis was used to demonstrate the main challenges of the interaction between students and AI. Although the AI chatbot provides more content-focused feedback (58.8%) than form-focused feedback (41.2%), the participants rejected more content than form-focused feedback. It suggests that they are less prepared to engage with content-focused than form-focused feedback. A comparison of rejected AI feedback between form-focused and content-focused revealed a similar pattern, with "Mismatch with learners' expectations" and "Anticipated high workload to respond" being rated as the top two reasons for rejection. As shown in Table 4, "Mismatch with external references" is mentioned

Table 4. Descriptive data and proportion of the students' rejection.

Factors of UTAUT	Form—Rejection		Content—Rejection	
Mismatch with learners' expectations	25	36.2%	37	37.8%
Anticipated high workload to respond	24	34.8%	33	33.7%
Impeding conditions	16	23.2%	7	7.1%
Mismatch with external references	4	5.8%	21	21.4%
Sum of rejection	69		98	

less frequently to reject form-focused AI feedback, while "Impeding conditions" is reported less frequently to reject content-focused feedback.

Ordinal logistic regression was conducted to determine which factors were statistically significant in predicting learners' rejection of AI feedback after controlling pre-experiment academic writing proficiency. Regardless of the focus of the feedback, the chi-square test shows that the assumption of parallelism tested was fulfilled, including Form-focused ($\chi^2 = .000$, $p > .05$) and Content-focused ($\chi^2 = .000$, $p > .05$) feedback. As shown in Table 5, the model fitness ($\chi^2 = 129.912$, $p < .05$) showed a relationship between Form-focused feedback rejected by students and the challenges from the interaction. An increase in the "Anticipated high workload to respond" feedback results in a faster decline in the overall number of students rejecting feedback, which is the only factor statistically significant (Wald $\chi^2(1) = 2.432$, $p < .05$). In other words, students are more likely to reject form-focused feedback (such as formatting and grammatical errors) when they are cognitively overloaded by the overwhelming amount of feedback.

As for the Content-focused AI feedback, the model fitness ($\chi^2 = 151.836$, $p < .05$) showed a relationship between Content-focused feedback rejected by students and the challenges from the interaction. Notably, all the main factors in the model had statistical significance, as demonstrated in Table 6. The feedback provided by AI had a greater impact on students' rejection in terms of "Mismatch with learners' expectations" (Wald $\chi^2(1) = 13.677$, $p < .000$), "Anticipated high workload to respond" (Wald $\chi^2(1) = 18.253$, $p < .000$) and "Mismatch with external references" (Wald $\chi^2(1) = 14.541$, $p < .000$), when compared to "Impeding conditions" (Wald $\chi^2(1) = 6.877$, $p < .000$). In summary, it may be interpreted from four perspectives. Firstly, AI-generated content-focused feedback may not align with participants' expectations or be perceived as useful, creating a disconnect between the AI's output and the student's learning objectives. Secondly, the level of effort required to engage with AI feedback can be significantly higher than what students are able to accommodate, leading to cognitive overload and disengagement. Thirdly, AI-generated feedback may not correspond to or support the feedback provided by teachers and peers, causing confusion and inconsistency in the learning process. Lastly, students may lack the necessary skills to comprehend and respond to AI-generated content-focused feedback, further exacerbating the issue.

Table 5. Reasons for rejecting AI-generated form-focused feedback.

Effect	Estimate	SE	95% CI		p
			LL	UL	
Mismatch with learners' expectations	−33.339	21.387	−75.239	8.562	.119
Anticipated high workload to respond	−33.332	13.623	−60.032	−6.632	.014
Mismatch with external references	−11.137	7.336	−25.516	3.241	.129
Impeding conditions	−22.414	11.948	−45.832	1.003	.061
Model fit (−2 log-likelihood)					.000

Table 6. Reasons for rejecting AI-generated content-focused feedback.

Effect	Estimate	SE	95% CI		p
			LL	UL	
Mismatch with learners' expectations	−32.096	8.679	−49.106	−15.086	.000
Anticipated high workload to respond	−23.059	5.397	−33.638	−12.481	.000
Mismatch with external references	−25.894	6.790	−39.203	−12.585	.000
Impeding conditions	−7.433	2.834	−12.988	−1.878	.009
Model fit (−2 log-likelihood)					.000

Discussion

One of the objectives of this study was to evaluate the primary challenges encountered by L2 learners when they receive feedback generated by ChatGPT. Following the UTAUT, we identified four major obstacles that could arise during ChatGPT interaction: "Mismatch with learners' expectations," "Anticipated high workload to respond," "Mismatch with external references," and "Impeding conditions."

To address the first research question, the current research found that students tend to reject AI feedback due to the low quality of feedback provided by AI and the significant amount of time and effort required to learn and understand AI-generated feedback. A possible explanation for subpar AI feedback could be attributed to the inherent limitations of the technology. This aligns with the findings from Winkler et al.'s (2020) research, which proposed that the chatbot's inadequacy in providing accurate and fitting feedback may stem from its restricted datasets. Another reason for students' rejection appears to align with the study conducted by Ericsson et al. (2023), which indicated that students who were required to spend significant amount of time to adapt to a new technology might experience a loss of confidence over time, resulting in a reluctance to use the technology continuously. At the same

time, according to statistical analysis, peer and teacher encouragement and support have a limited impact. Interestingly, this finding differs somewhat from previous research, which suggested that learners' high-level utilization of chatbots with the conditions aligned with assignment and learning tasks (Hew et al., 2023). The observed discrepancy can potentially be attributed to the fact that students prioritize self-evaluation on AI feedback over external evaluations such as teachers or other evaluators.

Another aim of this study was to examine the impact of various factors on students' negative behaviors with ChatGPT, with a specific focus on the feedback types related to form and content. Concerning form-focused feedback rejected by students, it was found that only the "Anticipated high workload to respond" feedback significantly affected the rejection of AI feedback. Upon conducting a text analysis, we identified two significant challenges students face when receiving ChatGPT-generated feedback. The first challenge is the issue of overwhelming information in their feedback, making it difficult for students to absorb and comprehend. The second challenge is the tendency for feedback to be too general rather than highlighting specific mistakes. These findings were also reported by Yang et al. (2023), who indicated that feedback lacking examples and contextual information receives lower student utilization rates than error-corrective and general feedback. Several factors could explain this observation. Firstly, students prefer a passive learning approach, commonly practiced in traditional classrooms, as it appears less time-consuming and effort-intensive than active learning. Another possible explanation is that students may not possess adequate feedback literacy to understand and act upon external feedback. This is also supported by Carless and Boud (2018), who stated that seeking and applying feedback effectively are two essential features of learners with high feedback literacy.

Regarding the content-focused feedback rejected by students, it is surprising that all four types of feedback significantly affected the rejection of AI feedback. The ordinal regression result shows that the challenge of "Mismatch with learners' expectations" and "Mismatch with external references" has the most significant predictive power in determining the rejection rate of students. In contrast, "Impeding conditions" have a minimal effect. In other words, the rejection rate of AI-generated feedback is primarily influenced by perceived usefulness and its degree of alignment with teacher and peer feedback. AI-generated content-focused feedback that does not align with the suggestions of teachers and peers and is perceived as less useful in improving the quality of assignments is less likely to be accepted and responded to positively. These results support earlier research that examined the link between an individual's self-perception of usefulness and their likelihood to use technology, highlighting that the perceived usefulness of technology is a significant predictor of behavioral intention (Chen et al., 2020). However, the negative emotions and feelings students experience while seeking feedback and learning with an AI chatbot are not the primary factors contributing to their rejection of ChatGPT-generated feedback. This interesting result could be due to students' valuing the utility of the feedback over their emotional response to it.

Limitations

Despite the valuable findings of this research, several limitations must be acknowledged. Firstly, while our study provided an analysis of feedback, a more systematic and in-depth analysis could have more significant pedagogical implications for both language teachers and learners. The scope of our analysis was limited, which only targeted mistakes and errors at two levels: form-focused and content-focused. A more comprehensive examination of feedback, considering various aspects such as tone, clarity, specificity, and relevance, could provide more nuanced insights into the efficacy of AI-generated feedback. Secondly, the sample size of our study was relatively small, and the participants were predominantly from higher education. This specific demographic might only be representative of some L2 learners, potentially limiting the generalizability of our findings to other learner groups. Thirdly, although our study focuses on learners' engagement strategies, it is worth noting that data triangulation, incorporating multiple data sources (such as learners' self-reported data), could enhance the reliability and validity of our findings. This approach could allow future studies to generalize the findings on a larger scale and provide a more comprehensive understanding of L2 learners' experiences with AI-generated feedback. In conclusion, future research should address these limitations and further explore this complex issue.

Conclusion

Guided by the UTAUT model, our research aimed to investigate the challenges encountered by L2 learners when seeking corrective feedback on academic writing from ChatGPT. Initially, this study identifies four significant challenges faced by L2 learners when receiving feedback generated by ChatGPT, encompassing factors such as "Mismatch with learners' expectations," "Anticipated high workload to respond," "Mismatch with external references," and "Impeding conditions." Subsequently, our investigation delves into the impact of the four barriers on students' rejection of form-focused and content-focused feedback. Concerning form-focused feedback, we find that the main significant predictor of student rejection is the extraordinarily high level of effort required to comprehend feedback. Conversely, the rejection rate of AI-generated content-focused feedback is influenced by various factors, including the disparity between students' expectations and the actual feedback received, feedback necessitating substantial comprehension time, AI-generated feedback conflicting with teacher suggestions, and individual limitations in cognitive and emotional abilities. These findings offer valuable insights into the challenges L2 learners encounter in accepting and utilizing AI-generated feedback.

In response to our study's findings on the challenges L2 learners face when using ChatGPT for feedback in academic writing, we propose specific pedagogical strategies. We recommend

integrating AI feedback with reliable external learning resources to mitigate inconsistencies and developing comprehensive support materials to assist students in comprehending formative corrective AI comments. To address the significant effort required to process feedback, we suggest establishing courses on computational thinking and designing materials to enhance critical evaluation skills (Ng et al., 2022). These strategies are intended to develop students' ability to interpret and utilize AI-generated feedback, thus enhancing their academic writing.

Contributorship

Ziqi Chen was responsible for analyzing the data, composing both the Abstract and the bulk of the main body, and addressing the reviewers' comments. Wei Wei focused on compiling the recent literature regarding the impact of technology-assisted feedback on second language teaching and learning, contributing significantly to the paper's Literature Review and Discussion sections. Xinhua Zhu and Wei Wei examined various aspects of the dissertation, including theoretical analyses of the challenges students encounter while revising their essays with feedback generated by ChatGPT, and how these challenges influence students' rejection of ChatGPT-generated feedback. Yuan Yao was responsible for finalizing the thesis and offered valuable guidance throughout the writing process.

Declaration of conflicting interests

Ethical statement

Funding

ORCID iDs

Ziqi Chen   https://orcid.org/0000-0003-3139-9723
Yuan Yao   https://orcid.org/0000-0003-0665-7065

References

Bećirović, S. (2023). Challenges and barriers for effective integration of technologies into teaching and learning. In S. Bećirović (Ed.), *Digital pedagogy* (pp. 123–133). Springer. https://doi.org/10.1007/978-981-99-0444-0_10

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Celik, I. (2023). Exploring the determinants of artificial intelligence (AI) literacy: Digital divide, computational thinking, cognitive absorption. *Telematics and Informatics*, *83*(1), 102026. https://doi.org/https://doi.org/10.1016/j.tele.2023.102026

Chen, H.-L., Vicki Widarso, G., & Sutrisno, H. (2020). A ChatBot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, *58*(6), 1161–1189. https://doi.org/10.1177/0735633120929622

Cheong, C. M., Luo, N., Zhu, X., Lu, Q., & Wei, W. (2023). Self-assessment complements peer assessment for undergraduate students in an academic writing task. *Assessment & Evaluation in Higher Education*, *48*(1), 135–148. https://doi.org/10.1080/02602938.2022.2069225

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Ericsson, E., Sofkova Hashemi, S., & Lundin, J. (2023). Fun and frustrating: Students' perspectives on practising speaking English with virtual humans. *Cogent Education*, *10*(1), 2170088. https://doi.org/10.1080/2331186X.2023.2170088

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*(1), 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*(1), 100775. https://doi.org/10.1016/j.jslw.2020.100775

Fitria, T. N. (2021). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. *Metathesis: Journal of English Language, Literature, and Teaching*, *5*(1), 65–78. https://doi.org/10.31002/metathesis.v5i1.3519

Fryer, L. K., Thompson, A., Nakao, K., Howarth, M., & Gallacher, A. (2020). Supporting self-efficacy beliefs and interest as educational inputs and outcomes: Framing AI and human partnered task experiences. *Learning and Individual Differences*, *80*(1), 101850. https://doi.org/10.1016/j.lindif.2020.101850

Gallacher, A., Thompson, A., & Howarth, M. (2018). "My Robot is an Idiot!"—Students' perceptions of AI in the L2 classroom. Research presentation at EUROCALL International Convention, Jyväskylä, Finland. https://doi.org/10.14705/rpnet.2018.26.815

Ghufron, M. A., & Rosyida, F. (2018). The role of Grammarly in assessing English as a foreign language (EFL) writing. *Lingua Cultura*, *12*(4), 395–403. https://doi.org/10.21512/lc.v12i4.4582

Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *The Journal of Information and Systems in Education*, *13*(1), 1–7. https://doi.org/10.12937/ejsise.13.1

Guo, K., Li, Y., Li, Y., & Chu, S. (2023). Understanding EFL students' chatbot-assisted argumentative writing: An activity theory perspective. *Education and Information Technologies*, *29*(1), 1–20. https://doi.org/10.1007/s10639-023-12230-5

Hatlevik, O. E., Throndsen, I., Loi, M., & Gudmundsdottir, G. B. (2018). Students' ICT self-efficacy and computer and information literacy: Determinants and relationships. *Computers & Education*, *118*(1), 107–119. https://doi.org/10.1016/j.compedu.2017.11.011

Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education*, *35*(1), 40–68. https://doi.org/10.1007/s12528-022-09338-x

Hinostroza, J., Ibieta, A., Claro, M., & Labbé, C. (2015). Characterisation of teachers' use of computers and internet inside and outside the classroom: The need to focus on the quality. *Education and Information Technologies*, *21*(6), 1595–1610. https://doi.org/10.1007/s10639-015-9404-6

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Logistic regression models for multinomial and ordinal outcomes. In *Wiley Series in Probability and Statistics* (pp. 269–311). https://doi.org/10.1002/9781118548387.ch8

Howard, S. K. (2013). Risk-aversion: Understanding teachers' resistance to technology integration. *Technology, Pedagogy and Education*, *22*(3), 357–372. https://doi.org/10.1080/1475939X.2013.802995

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. *Languages*, *8*(4), 247. https://doi.org/10.3390/languages8040247

Kaebnick, G. E., Magnus, D., Kao, A., Hosseini, M., Resnik, D. B., Dubljević, V., Rentmeester, C. A., Gordijn, B., & Cherry, M. J. (2023). Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing. *Ajob Neuroscience*, *14*(4), 337–340. https://doi.org/10.1080/21507740.2023.2257181

Kaur, A., Singh, S., Chandan, J. S., Robbins, T., & Patel, V. (2021). Qualitative exploration of digital chatbot use in medical education: A pilot study. *Digit Health*, *7*(2021), 20552076211038151. https://doi.org/10.1177/20552076211038151

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Lai, C. (2015). Modeling teachers' influence on learners' self-directed use of technology for language learning outside the classroom. *Computers & Education*, *82*(1), 74–83. https://doi.org/10.1016/j.compedu.2014.11.005

Leem, J., & Sung, E. (2019). Teachers' beliefs and technology acceptance concerning smart mobile devices for SMART education in South Korea. *British Journal of Educational Technology*, *50*(2), 601–613. https://doi.org/10.1111/bjet.12612

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment & Evaluation in Higher Education*, *49*(5), 616–633. https://doi.org/10.1080/02602938.2024.2301722

Marchante, B. M. (2022). The use of ICTs and artificial intelligence in the revision of the writing process in Valencian public universities. *Research in Education and Learning Innovation Archives*, *28*(2022), 16. https://doi.org/10.7203/realia.28.20622

McDonald, N., Johri, A., Ali, A., & Hingle, A. (2024). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2402.01659

Mohebbi, H. (2021). 25 years on, the written error correction debate continues: An interview with John Truscott. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(1), 3. https://doi.org/10.1186/s40862-021-00110-9

Morales, L., & Roig, G. (2002). Connecting a technology faculty development program with student learning. *Campus-Wide Information Systems*, *19*(2), 67–72. https://doi.org/10.1108/10650740210421881

Ng, D. T. K., Luo, W. Y., Chan, H. M. Y., & Chu, S. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. *Computers & Education: Artificial Intelligence*, *3*(1), 100054. https://doi.org/10.1016/j.caeai.2022.100054

Reiser, L. J. (2002). Professional development and other factors that contribute to the ability to integrate technology into curriculum. *Journal of Educational Technology Systems*, *30*(4), 437–460. https://doi.org/10.2190/jb9u-gjbd-lrur-0vwu

Rozencwajg, S., & Kantor, E. (2023). Elevating scientific writing with ChatGPT: A guide for reviewers, editors … and authors [editorial]. *Anaesthesia Critical Care & Pain Medicine*, *42*(3), 101209. https://doi.org/10.1016/j.accpm.2023.101209

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care*, *27*(1), 75. https://doi.org/10.1186/s13054-023-04380-2

Shin, W. S., Han, I., & Kim, I. (2014). Teachers' technology use and the change of their pedagogical beliefs in Korean educational context. *International Education Studies*, *7*(8), 11–22. https://doi.org/10.5539/ies.v7n8p11

Su, J., Ng, D. T. K., & Chu, S. K. W. (2023). Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. *Computers and Education: Artificial Intelligence*, *4*(1), 100124. https://doi.org/10.1016/j.caeai.2023.100124

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478. https://doi.org/10.2307/30036540

Wei, W., Cheong, C. M., Zhu, X., & Lu, Q. (2022). Comparing self-reflection and peer feedback practices in an academic writing task: A student self-efficacy perspective. *Teaching in Higher Education*, *29*(4), 896–912. https://doi.org/10.1080/13562517.2022.2042242

Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, USA. https://doi-org.libezproxy.um.edu.mo/10.1145/3313831.3376781

Yang, H., Gao, C., & Shen, H.-z. (2023). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies*, *29*(4), 3837–3858. https://doi.org/10.1007/s10639-023-11991-3

Yang, H.-C., & Zapata-Rivera, D. (2010). Interlanguage pragmatics with a pedagogical agent: The request game. *Computer Assisted Language Learning*, *23*(5), 395–412. https://doi.org/10.1080/09588221.2010.520274

Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing*, *43*(1), 100439. https://doi.org/10.1016/j.asw.2019.100439

Zimmerman, A. (2023). A ghostwriter for the masses: ChatGPT and the future of writing. *Annals of Surgical Oncology*, *30*(6), 3170–3173. https://doi.org/10.1245/s10434-023-13436-0