Feedback seeking abilities of L2 writers using ChatGPT: a mixed method multiple case study

Abstract

Purpose: The study investigated the feedback seeking abilities of learners in L2 writing classrooms using ChatGPT as an automated written corrective feedback (AWCF) provider. Specifically, the research embarked on the exploration of L2 writers' feedback seeking abilities in interacting with ChatGPT for feedback and their perceptions thereof in the new learning environment.

Design/methodology/approach: Three EFL learners of distinct language proficiencies and technological competences were recruited to participate in the mixed method multiple case study. The researcher used observation and in-depth interview to collect the ChatGPT prompts written by the participants and their reflections of feedback seeking in the project.

Findings: The study revealed that: 1) students with different academic profiles display varied abilities to utilize the feedback seeking strategies; 2) the significance of feedback seeking agency was agreed upon; and 3) the promoting factors for the development of students' feedback seeking abilities are the proactivity of involvement and the command of metacognitive regulatory skills.

Implications: Additionally, a conceptual model of feedback seeking in an AI-mediated learning environment was postulated. The research has its conceptual and practical implications for researchers and educators expecting to incorporate ChatGPT in teaching and learning. The research unveiled the significance and potential of using state-of-the-art technologies in education. However, since we are still in an early phase applying such tools in authentic pedagogical environments, many instructional redevelopment and rearrangement should be considered and implemented.

Originality/value: The work is a pioneering effort to explore learners' feedback seeking abilities in a ChatGPT-enhanced learning environment. It pointed out new directions for process-, and student-oriented research in the era of changes.

Keywords: ChatGPT; AI-enhanced learning; feedback seeking abilities; mixed method case study; corrective feedback

Introduction

In educational studies, feedback has been understood as one of the most effective and easy-to-implement instructional interventions. Researchers have agreed upon the positive

effects and critical position of feedback in enhancing learning achievement (e.g., Hattie and Timperley, 2007; Wiliam, 2018). Similarly, Hyland and Hyland (2019) point out that feedback remains a central topic for educators and researchers in the field of second language (L2) writing with its affirmed role in "encouraging and consolidating learning" (p. 1). In light of the technological advancements, a research trend to study the impact of automated written corrective feedback (AWCF) has become increasingly visible in recent years (Barrot, 2021).

Since the advent of ChatGPT, a conversational service based on generative artificial intelligence (GAI), researchers have enthusiastically explored its capacity for an array of disciplines. According to several piloting studies, ChatGPT has the potential to outperform and replace its predecessors as an error corrector or AWCF provider (Wu, Wang, *et al.*, 2023; Yan, 2023). ChatGPT's power to automate tasks relies on *prompt*s, the interactive instructions created by users inclusive of specific job demands, contextual backgrounds, objectives for the automation, etc. Consequently, L2 writers are expected to proactively *seek* instead of *receive* feedback from the system with well-crafted prompts (White *et al.*, 2023). In a conventional L2 setting, the concept of feedback seeking behavior (FSB) facilitates the understanding of learners' inquiry and monitoring of received feedback from teachers or peers (Papi *et al.*, 2019). However, positioned in the new and interactive learning environment, we are facing the scantness of empirical evidence in students' actual abilities to seek feedback generated by AI-based systems.

Against the above backdrops, the study sets out to explore learners' feedback seeking abilities in L2 classrooms using ChatGPT. Three participants with different levels of language proficiencies and technological competences were recruited for the multiple case studies employing the sequential explanatory mixed-method design. After a 5-week project, the researcher analyzed ChatGPT prompts created by the participants through

coding and quantitative document analysis. Then, transcripts of the two in-depth interview sessions were analyzed for qualitative findings. Through the data triangulation and interpretation, the following research questions are answered:

RQ1: What characterizes the feedback seeking abilities of the participants in the interaction with ChatGPT for feedback?

RQ2: How do L2 writers perceive their feedback seeking abilities in an ChatGPT-enhanced learning environment?

The study is significant as an innovative effort to explore the learning behaviors of students in a brand-new environment. The conceptual and practical contribution directed from the findings would be valuable for the field of AI-mediated language education. The next section provides a review of significant literature related to the core concept and theories related to the topic. Section 3 details the methodologies utilized in the study. Results and the interpretation are presented in Section 4 and 5 respectively before the concluding remarks, limitations, and future research directions in Section 6.

Literature Review

In this section, literature on feedback seeking behavior and automated feedback providers is reviewed before the conceptualization of feedback seeking in a GAI-enhanced learning environment.

Feedback seeking behavior

FSB, the processes of learners' seeking feedback information, is one of the determining factors of feedback effectiveness (Leenknecht *et al.*, 2019). The dyadic construct, composing of feedback *inquiry* and *monitoring*, was borrowed from Ashford et al. (2003) to indicate the purposeful acquisition of feedback information from the

surrounding environment or agents. Papi et al. (2019) recast the term for language learning to denote students' utilization of inquiry and monitoring strategies in seeking feedback from teachers, peer learners and the learning environment.

The introduction of FSB to language learning reflects: 1) the focus on individual differences among learners in feedback (Winstone *et al.*, 2021); 2) the shift from viewing feedback as a teaching to a learning resource (Papi *et al.*, 2019); and 3) the emphasize on the proactive *involvement* instead of the *reception* of learners (Malecka *et al.*, 2022). Two research strands could be identified from existing literature on FSB in L2 learning. On the one hand, extant studies focused on the relation between FSB, its motivational antecedents, and/or subsequent learning outcomes (see Xu and Wang, 2023; Yao and Zhu, 2022). On the other hand, a few recent studies have delved into L2 learners' FSB in technology-enhanced settings. For example, Kessler (2023) argued that L2 learners in an online community actively sought feedback through *clarification* and *confirmation* requests with clear individual strategic tendencies. Similarly, Xu (2021) studied L2 writers' FSB during Covid-19 and discovered the positive correlation between feedback-seeking orientations and the use of self-regulatory strategies.

However, we still face a paucity of knowledge in how students seek feedback in emerging environments practicing *new feedback paradigm*, where the role of teachers is minimal and students' information-generation from and sense-making of technology-produced feedback are pivotal (Jensen *et al.*, 2023). FSB from existing studies remains of the *old paradigm*, in which the students, receivers of feedback, manage to strike a balance between the explicit inquiry for feedback from teachers and peers, and the implicit monitoring of the learning environment (Papi *et al.*, 2019). Conversely, the changing learning environment as well as the role of learners as feedback *generators* urge further exploration into students' feedback seeking in ChatGPT-enhanced settings.

Automated feedback providers: potential of ChatGPT

Researchers have contributed a plethora of studies exploring the effects of technology-enhanced feedback on language learning, especially L2 writing performance. In the beginning, corpus-based automated writing evaluation (AWE) systems such as Pigai, iWrite, and iTEST were widely applied in L2 writing classrooms (Zhang, 2021). Soon the AWE systems gave place to AI-based corrective feedback providers, e.g., Grammarly and QuillBot. The advantage of these AI-based solutions were empirically validated (Ngo *et al.*, 2022) and thus remained the most frequent choices for automated feedback in L2 classrooms.

Since its debut, researchers have attempted to investigate into ChatGPT's effects on L2 writing. Yan (2023) explored L2 writers' abilities to utilize ChatGPT's text generation functionality in writing and their perceptions towards the potential threats to academic integrity. Barrot (2023) summarized advantages and pitfalls of ChatGPT's potential usage in L2 writing classrooms, and offered recommendations for L2 teachers. Furthermore, efforts have been made to incorporate ChatGPT in L2 writing assessment. Mizumoto and Eguchi (2023) reported convincing accuracy and reliability from the automated scoring on L2 writing products by ChatGPT. Accordingly, the potential of ChatGPT as a powerful AWCF providers is posited on: 1) the ability to provide accurate grammatical and syntactical corrections, thanks to the tremendous volume of trained linguistic data (Fang *et al.*, 2023); 2) the mechanism of prompt engineering with which users could tap prompts to iteratively and progressively elicit feedback and responses of improved quality (White *et al.*, 2023); and 3) the observed enhancement of learners' engagement and involvement in the interaction with ChatGPT (Yan, 2023).

Furthermore, ChatGPT brings changes to L2 learning behaviors in seeking and utilizing feedback. First, the human-ChatGPT communication is bi-directional, making ChatGPT a learner-aware, human-like, and adaptive feedback provider (Barrot, 2023). Second, once properly prompted, ChatGPT could generate feedback on a wide range of writing constructs and linguistic features, inclusive but not limited to grammar and syntactical correctness (Mizumoto and Eguchi, 2023). Third, multi-turn human-ChatGPT communication is encouraged and corroborated to be conducive to performance improvement of ChatGPT-generated feedback (Bang *et al.*, 2023). Given the human-AI communication interactivity and corresponding changes in learning behaviors, relevant studies on learning behavior from the learner perspectives are urgently needed.

Conceptualizing feedback seeking in GAI-enhanced learning

The study is grounded on interaction theory. As argued by Long (1996), the four constructs involved in learning included input, interaction, feedback and output. For L2 development, Bitchener and Ferris (2012) emphasized the understanding, internalization and processing of feedback on the basis of exiting knowledge. Furthermore, Kaptelinin and Nardi (2018) pinpointed that the deepening of the effects of technology could be attained by putting human-computer interaction within the "purposeful human activities" (p. 3). Correspondingly, a conceptual model (**Figure 1**) is proposed by integrating feedback seeking from GAI-based systems with the subsequent internalization and processing of the feedback.

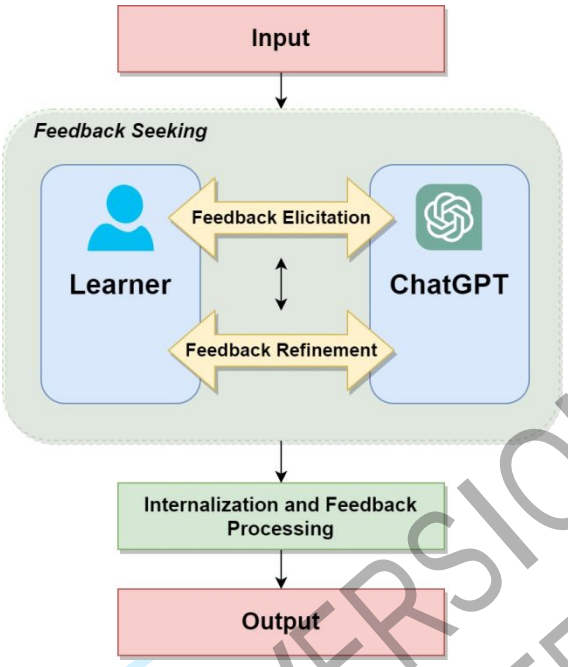


Figure 1. Conceptual Model of L2 writers' feedback seeking from ChatGPT

According to Papi et al. (2019), FSB consists of feedback monitoring and inquiry through which the learners observe and absorb relevant feedback information from the environment, and seek further explanation of the feedback. However, the emergence of GAI-based tools poses new challenges to the operationalization of the two behaviors. First, the approaches of feedback inquiry changes. Compared to the traditional teacher-fronted settings, learners no longer resort to the explanation and help from teachers or peer (Jensen *et al.*, 2023). Conversely, the affordance of GAI systems could replicate the role of both teachers and peer learners at the cost of extra prompts written and submitted through the dialog interfaces. Second, the feedback monitoring as described by Papi et al. (2019) is irrelevant for the setting with ChatGPT. With the conversational interaction with AI and proper prompt writing techniques, learners could instantly and iteratively request GAI systems to amend, improve and regenerate feedback in a rather explicit and direct manner. Hence, during a feedback episode, the feedback seeking processes could be explicitly executed for multiple times. Consequently, the original two component of

FSB could be reframed: 1) feedback elicitation, in which the learners use prompts to iteratively generate feedback with different contents and focus; 2) feedback refinement, an on-going process in which the learners repeatedly report their inner judgement to the GAI systems, after which feedback would be regenerated based on the provided evaluation. Correspondingly, feedback seeking ability refers to the capability of students to effectively use feedback elicitation and refinement strategies in the interaction with GAI systems.

Methods and Materials

Design

The study adopted a mixed method multiple case study approach (Yin, 2013). In the fields of applied linguistics (Duff, 2010) and instructional feedback (e.g., Zhang and Hyland, 2023), case study has been widely applied as an established means to collect rich data of students' actual learning experiences. Additionally, Adams (2019) argued that the limited number of research subjects in case studies had its merits in unfolding learner experiences in using the feedback other than the feedback design.

A case was defined as a student's ability to effectively use feedback seeking strategies and their perceptions thereof. Internally, a sequential explanatory mixed method design was selected so that the quantitively displayed characteristics of students' feedback seeking abilities would be explained by their own reflection and perceptions (Creswell and Plano Clark, 2018). The study was collective in nature, since it allowed the researcher to generalize the findings across multiple cases to a broader context (Stake, 1995).

Participants

The study took place in an EFL program at a Chinese university. Students in the program were required to take two writing courses. The researcher purposefully sampled three participants with distinct level of language proficiencies and technological competences (see **The screening** criteria included: 1) mean grades in precedent L2 writing assessments; 2) experiences and interest in computer-assisted language learning and feedback reflected in a pre-project survey; 3) performance in two digital humanities courses; and 4) self-rated technological competences. The criteria ensured that students with various levels of language and technological competence were recruited, thus forming representative cases of L2 writers in the program.

Table I for their characteristics). The screening criteria included: 1) mean grades in precedent L2 writing assessments[1]; 2) experiences and interest in computer-assisted language learning and feedback reflected in a pre-project survey; 3) performance in two digital humanities courses; and 4) self-rated technological competences[2]. The criteria ensured that students with various levels of language and technological competence were recruited, thus forming representative cases of L2 writers in the program.

Table I. Demographic Information and Characteristics of the Participants

Pseudonyms	Gender	Age	Mean Grades 1*	Mean Grades 2*	Self-rated Technological Competences	Instructors' Recommendations
Toby	M	21	81/100	91/100	9/10	"His balance in both language proficiency and technology abilities make

[1] Writing tasks were developed as per the requirement of College English Test Band 6, an English proficiency examination commonly administered in Chinese universities. The writing performances were rated against a 100-mark rubric composing five writing assessment dimensions, i.e., content, organization, grammar, vocabulary, and mechanics.

[2] Measured by a self-reported 10-point survey including the following dimensions: a) general digital literacy; b) information retrieval and evaluation; c) communicative and collaborative skills using technology; and d) language-specific technological skills.

						him a must choice for your consideration"
Lisa	F	19	67/100	85/100	8/10	"Lisa is a rare case for female students, she is better at handling the technology-related stuffs".
Isabella	F	20	88/100	62/100	5/10	"Just a typical liberal arts student, at least a very hard-working student in the digital humanity courses ".

*Mean Grades 1: mean grades in precedent L2 writing assessments.
Mean Grades 2: mean grades in precedent assessments in digital humanities courses.

Based on students' performance in pre-project language assessments and teachers' recommendations, Toby represented those with high language proficiency and technological competence; Lisa stood for those with medium language proficiency and high technological competence; and Isabella exemplified those with high language proficiency and medium technological competence.

Data Collection

The project was a five-week teacher-scaffolded informal L2 writing practicum with a total number of 68 learners (including the three participants). During the project, all learners were required to complete and revise three argumentative essay each no less than 250 words. Additionally, two sessions of live demonstration and peer discussion were arranged for each week (see Appendix S1 for practicum details). To ensure students' engagement with the practicum, they were allowed to choose one of the essays completed in the practicum as assignment (accounting for 15% of the final grade) of an ongoing writing course.

During the project, students were asked to keep track of the prompts they have written to interact with ChatGPT in learning logs (see Appendix S2). To overcome the potential bias in the focus and scope of ChatGPT-generated feedback due to different feedback seeking abilities and ChatGPT prompting strategies, all the participants were

trained and required to use the same ChatGPT prompting template (see Appendix S1). The official GPT-3.5 version of ChatGPT from OpenAI was used in the study. Students were encouraged to explore ChatGPT's functionality at their own paces, hence no time restraint was posed on their interaction with ChatGPT. The learning log could be filled with multimedia files as supplementary documents, e.g., screenshot, videorecording, etc. Furthermore, to record all the learning behaviors, the project took place at the computerized language laboratories equipped with real-time screen recording, keylogging, and video-based classroom monitoring devices. Upon the completion of each session in the project, students are asked to review and member-check the contents of learning log.

To obtain students' perceptions of their feedback seeking abilities in the ChatGPT-enhanced learning environment, two sessions of face-to-face in-depth interview were arranged for each participant at the third and fifth week respectively. Each interview session lasted for approximately 30 minutes. The interview processes strictly followed the interview protocol (see Appendix S3). The participants used their first language (Chinese) so that they could better express their understandings and perceptions. All contents and procedures of the interview were audio recorded, translated into English, and transcribed verbatim. To enhance the trustworthiness of the collected data, all translated texts were checked by a co-researcher who was a certified and accredited English-Chinese translator before member-checked by the participants (Birt *et al.*, 2016).

Coding and Analysis

Data obtained from observation and interviews were analyzed and organized in accordance with the within-case and cross-case analysis techniques recommended by seminal works of case study (Stake, 2005; Yin, 2013).

First, quantitative document analysis was applied to the ChatGPT prompts. In practice, a coding schemes of ChatGPT prompt patterns developed based on the experiences from the pilot study and the prompt engineering mechanism (White *et al.*, 2023) was applied in the study (see Appendix S4). Four coders were recruited from the teaching faculty to code the patterns of students' prompts. The inter-rater reliability measured by the Fleiss' Kappa statistics (Kappa = .76; [95% CI: 0.68-0.81]) revealed an acceptable level of agreement between the coders. Then, descriptive analysis and visualization were applied to the coded data.

Second, thematic analysis was applied to the transcribed interview data. In specific, the six-step procedures advocated by Braun and Clarke (2006) was adhered to. Two co-researchers were recruited to assist the researcher in the data coding and analysis. Initial codes of significant and recurring patterns were extracted from the original data, then they were refined into higher-order themes representing students' perceptions of their feedback seeking abilities from ChatGPT in language learning. To ensure the trustworthiness of the qualitative analysis, researcher triangulation and peer debriefing were practiced.

Validity

Results

Quantitative Document Analysis

Toby (**Table II**) has created a total of 2,537 pieces of prompts in 167 rounds of interaction with ChatGPT using 850 minutes; Lisa **(Table III**) has written 1,648 pieces of prompts in 193 rounds of interaction with ChatGPT using 436 minutes; Isabella (**Table IV**) has spent 1,034 minutes in her 70-round interaction with ChatGPT with 1,921 pieces of

prompts.

Table II. ChatGPT Prompts Created by Toby

	Prompt	Words	Words per prompt	Rounds	Prompts per round	Time spent (min)
Week 1	306	3213	10.5	48	6.4	251
Week 2	426	4856	11.4	35	12.2	215
Week 3	520	7124	13.7	31	16.8	177
Week 4	589	9365	15.9	27	21.8	111
Week 5	696	12389	17.8	26	26.8	96
Sum	2537	36947	-	167	-	850
Mean	507.4	7389.5	13.9	33.4	16.8	170

Table III. ChatGPT Prompts Created by Lisa

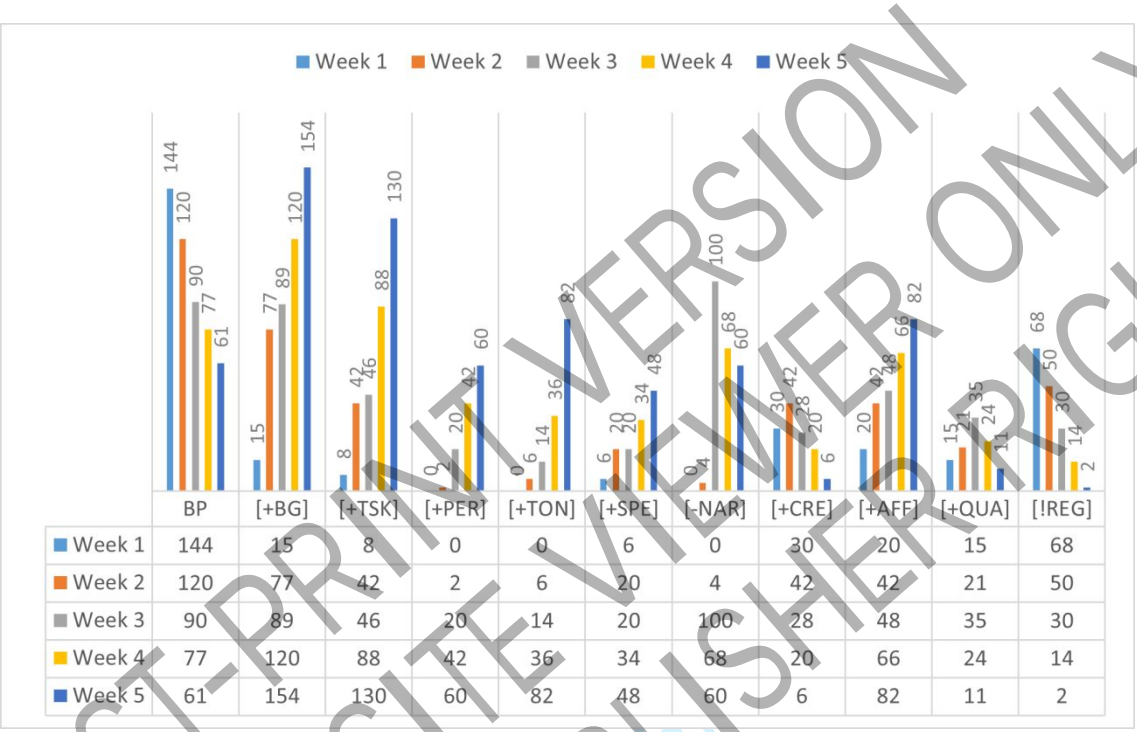
	Prompts	Words	Words per Prompt	Rounds	Prompts per Round	Time spent (min)
Week 1	249	2615	10.5	37	6.7	110
Week 2	292	3329	11.4	40	7.3	102
Week 3	301	3130	10.4	36	8.4	88
Week 4	375	4163	11.1	39	9.6	71
Week 5	431	4870	11.3	41	10.5	65
Sum	1648	18107	-	193	-	436
Mean	329.6	3621.3	10.9	38.6	8.5	87

Table IV. ChatGPT Prompts Created by Isabella

	Prompts	Words	Words per Prompt	Rounds	Prompts per Round	Time spent (min)
Week 1	245	3185	13.0	8	30.6	280
Week 2	335	4690	14.0	12	27.9	245
Week 3	442	7249	16.4	14	31.6	211
Week 4	432	6696	15.5	17	25.4	177
Week 5	467	9247	19.8	19	24.6	121
Sum	1921	31066	-	70	-	1034
Mean	384.2	6213.3	15.7	14	28	207

Against the coding scheme, the coders have categorized all prompts into two major dimensions of feedback elicitation and refinement. Then the data were further divided into 11 different subcategories of ChatGPT prompt patterns. With the coded data visualized by the temporal order, the development feature of the three participants' feedback seeking abilities were illustrated.

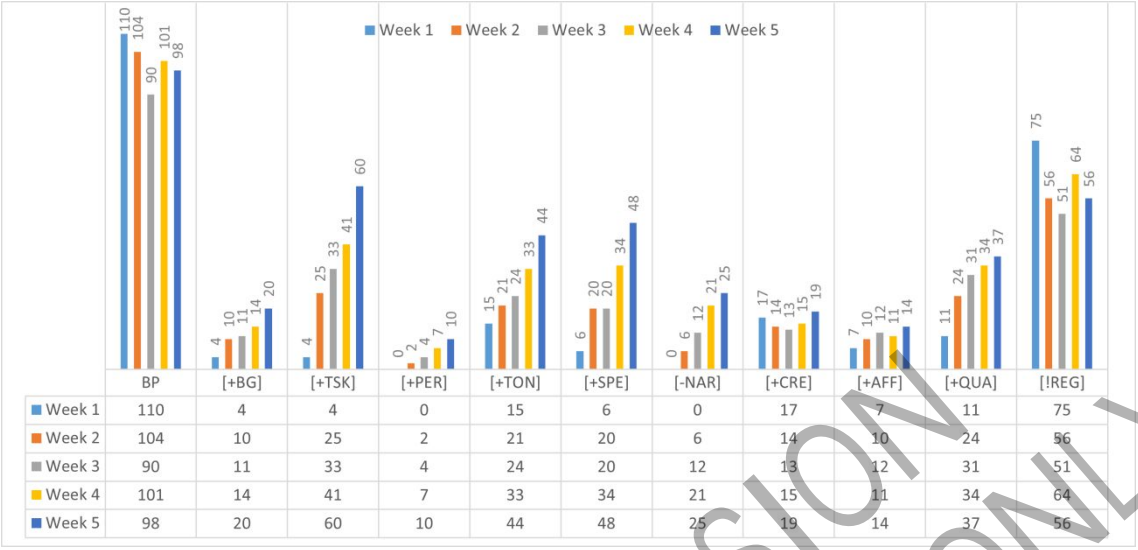
Toby's development of feedback seeking abilities was steady and balanced (see Figure 2). His application of basic prompts and direct feedback regeneration was remarkably reduced as he progressed; his utilization of other feedback elicitation strategies grew consistently while his command of feedback refinement strategies experienced a transition from a rapid heating up to a gradual cooling down.



	BP	[+BG]	[+TSK]	[+PER]	[+TON]	[+SPE]	[-NAR]	[+CRE]	[+AFF]	[+QUA]	[!REG]
Week 1	144	15	8	0	0	6	0	30	20	15	68
Week 2	120	77	42	2	6	20	4	42	42	21	50
Week 3	90	89	46	20	14	20	100	28	48	35	30
Week 4	77	120	88	42	36	34	68	20	66	24	14
Week 5	61	154	130	60	82	48	60	6	82	11	2

Figure 2. Prompt Patterns of Toby

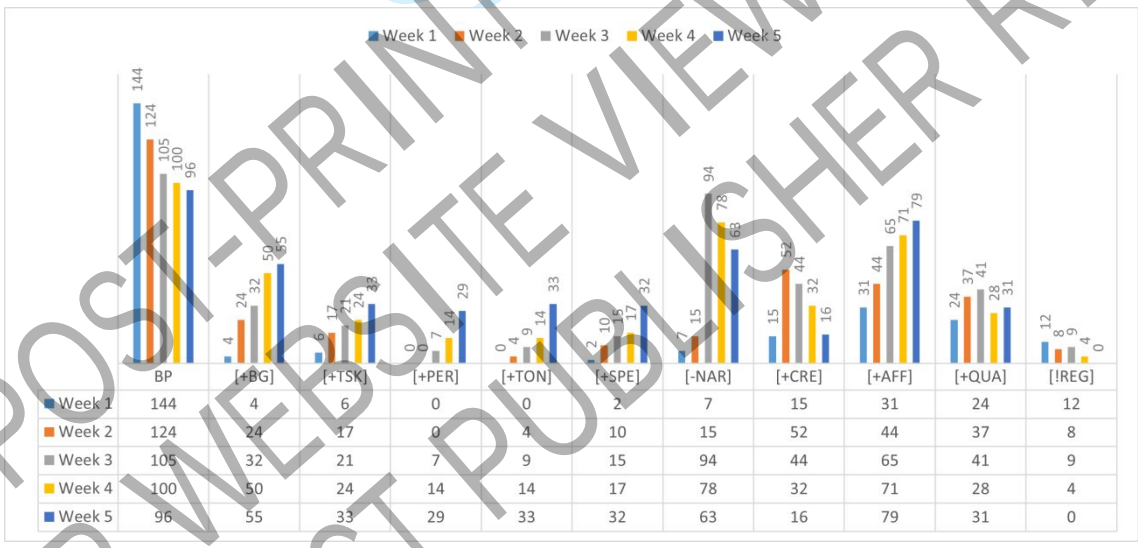
For Lisa (see Figure 3), her application of basic prompts and direct feedback regeneration remained generally unchanged in the project, while a growing momentum could be observed for the rest feedback seeking strategies.



	BP	[+BG]	[+TSK]	[+PER]	[+TON]	[+SPE]	[-NAR]	[+CRE]	[+AFF]	[+QUA]	[!REG]
Week 1	110	4	4	0	15	6	0	17	7	11	75
Week 2	104	10	25	2	21	20	6	14	10	24	56
Week 3	90	11	33	4	24	20	12	13	12	31	51
Week 4	101	14	41	7	33	34	21	15	11	34	64
Week 5	98	20	60	10	44	48	25	19	14	37	56

Figure 3. Prompt Patterns of Lisa

Finally, Isabella's development was similar to Toby's (see Figure 4), only relatively less in quantity as she appeared as a slow-paced and detailed prompt writer.



	BP	[+BG]	[+TSK]	[+PER]	[+TON]	[+SPE]	[-NAR]	[+CRE]	[+AFF]	[+QUA]	[!REG]
Week 1	144	4	6	0	0	2	7	15	31	24	12
Week 2	124	24	17	0	4	10	15	52	44	37	8
Week 3	105	32	21	7	9	15	94	44	65	41	9
Week 4	100	50	24	14	14	17	78	32	71	28	4
Week 5	96	55	33	29	33	32	63	16	79	31	0

Figure 4. Prompt Patterns of Isabella

Thematic Analysis

Through the thematic analysis, the following main themes were identified: 1) perceptions of feedback seeking from ChatGPT; 2) significance of active feedback seeking; and 3) factors for the development of feedback seeking abilities. The thematic map of the interviews is shown in Figure 5. Additionally, each of the participants summarized their

learning experiences in the second interview session.

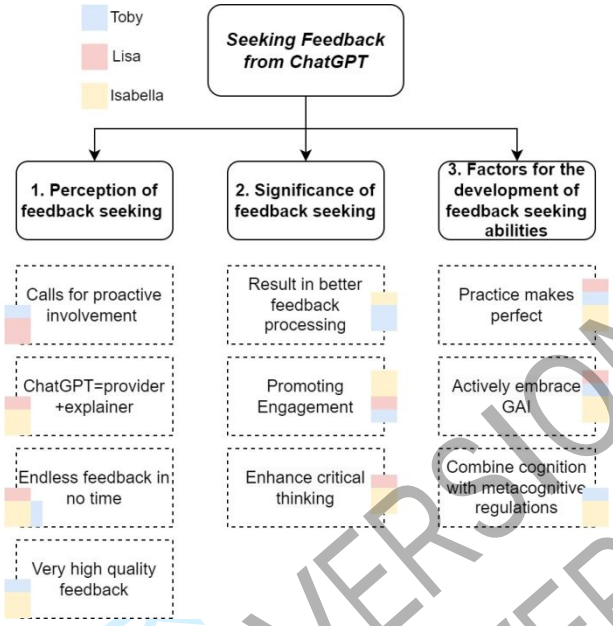


Figure 5. Thematic Map of the Interviews

Perception of feedback seeking from ChatGPT

When asked about the GAI-enhanced environments, Toby and Lisa stressed the proactive involvement of learners. As they have experienced, the shift from a feedback *receiver to seeker* was impactful on their "flow of learning". In the new environment, students' agency, creativity, and proactivity was highly valued.

Pertinent to the role ChatGPT played in L2 writing, Lisa and Isabella agreed upon the dual identity as both a feedback *provider* and *explainer*. From their viewpoints, ChatGPT availed the learners clearer and detailed explanation about the feedback. At the same time, they believed that such affordance distinguishes ChatGPT from previous feedback providers, as it could compensate for the possible unavailability of the teachers.

The three participants reached an agreement in the advantage of ChatGPT as an instant feedback provider. Learners could continuously solicit corrective feedback with improved quality from the AI system, if they were "well-versed in the language ChatGPT

spoke, aka, the prompts" (Toby, interview#1). Furthermore, Toby and Isabella argued in unison that "ChatGPT needs just a few seconds to provide feedback, and every second would be rewarded with corrections of high quality" (Isabella, interview#2). According to both Toby and Isabella, who have compared the quality of the corrections from multiple sources, e.g., Grammarly or QuillBot, ChatGPT was generally reliable and accurate.

Significance of active feedback seeking

All the three participants agreed that active feedback seeking was beneficial to the affective and behavioral outcomes of learning. Toby and Isabella emphasized that the subsequent processing of feedback gained from frequent and effective feedback seeking. Isabella underlined the value of the unfocused feedback offered by ChatGPT against the focused feedback from teachers that she "could make judgement on writing from a holistic perspective instead of paying attention merely to a certain aspect" (Isabella, interview#1).

Regarding the effects of the active feedback seeking from ChatGPT on the affective domain of learning experiences, all participants kept pace with each other. To them, the feedback seeking from ChatGPT helped creating a democratic atmosphere. Students were mentally at ease asking ChatGPT to regenerate or refine the feedback.

Furthermore, participants felt that they are motivated to actively involve in feedback seeking. In conventional settings, negative and critical feedback was treated by learners as shame. Consequently, students were generally reluctant to ask additional questions about the feedback, much less challenge it. The interactive processes to elicit feedback from ChatGPT, according to Lisa, "works like a black box, in which our shames are hidden, in a positive way" (Lisa, interview#2).

Finally, for the effects of feedback seeking on L2 writer's competence, Lisa and Isabelle pointed to students' critical thinking. In accordance with their reflections, the feedback refinement processes required the learners to critically compare the values and side effects of feedback under consideration. Furthermore, the abilities to accurately identify existing issues and critically describe them in writing prompts were vital for effective interaction with ChatGPT.

Factors for the development of feedback seeking abilities

Grounded on the consensus that ChatGPT provided a promoting environment for students to develop feedback seeking abilities, all participants accentuate that the experiences of interaction with AI system were most important. ChatGPT, from their viewpoints, is "shockingly easy-to-use at first glance" (Toby, interview#2), but takes "a huge amount of practice and thinking" to "craft the skills and techniques of writing and polishing prompts" (Isabella, interview#2).

Furthermore, all three participants agreed that an open-minded attitude mattered. According to Toby, the innovation ChatGPT brought about resulted in "either overhyped tunes or total rejections" (Toby, interview#1). As a result, there exists a gap between the perceived affordance of "ChatGPT as a key to good writing quality" (Lisa, interview#1) and realistic experiences that "it is a powerful yet demanding tool with diversified and unpredictable outcomes" (Isabella, interview#2). To mediate such cognitive dissonance, the participants suggested that deepening one's understanding of ChatGPT's mechanism and workflow in a real setting of L2 writing was imperative.

Toby and Isabella provided their take on the concrete learning strategies needed to foster feedback seeking abilities from AI-based systems. Reflected by their

experiences, learners' monitoring and evaluation were decisive for feedback quality. Toby confirmed that such refinements called for metacognitive abilities:

> *"To compare and select from multiple pieces of feedback, one needs to evaluate comprehensively, and use a whole flow of planning, comparing and controlling from a higher order perspective" (Toby, interview#2).*

Wrap-up of individual learning experiences

In the second interview sessions, each participant summarized their learning experiences and several post-project perceptions. In specific, their perceptions of feedback seeking, recommendations for the best practices and a summary of their personal growth were elucidated. The gist of their summaries is presented in **Table V**.

Table V. Individual Summary of Learning Experiences

Participant	**Toby**	**Lisa**	**Isabella**
Post-project perception of feedback seeking	A learners should actively seek feedback from AI systems. During the refinement, he/she should use multiple criteria to select and compare available feedback.	We should take advantage of AI system's ability to provide feedback from different angles.	Human understanding and judgement should not be replaced. One should think closely at the feedback provided by the AI system and make the final decision on his/her own.
Perceived best practice of feedback seeking	It's all about the balance. The balance between the learner's own judgement and the interaction with the system. Think hard for a clear decision, then interact proactively with the AI system to specify the details in the decision.	Proactively interact with ChatGPT for improvement in feedback. Learners should actively channel their own judgement back to the system. The more you interact with ChatGPT, the better.	Precision matters. When using the conversation with ChatGPT to produce feedback, we should give detailed and clear prompts. For the refinement, we should incorporate the AI's response with our own linguistics knowledge.
Personal growth	A thrilling experience in using ChatGPT for feedback. I like my fast-paced growth in feedback triggering and feedback monitoring. I have tried to be bold and exploratory in dealing with ChatGPT. I will explore more advanced tricks in the future, and I believe AI-based tools will be my learning companion in the days to come.	Have truly mastered the tricks to chat with AI. But I am not satisfied with my ability to write high quality prompt. I like the feeling with ChatGPT. But I will learn more about how to write effective prompt as a language learner.	I will describe my development as a slow-paced but fruitful one. Before the project, I am not confident in myself. So, for most of the time, I write prompts with measured words and am very careful talking with ChatGPT. But I like the feeling of reading and comparing the feedback.

Discussions

Being a multiple case study, the observations and findings emerged from individual cases were triangulated and holistically interpreted. The findings are discussed against existing empirical evidence and theoretical perspectives.

In answering **RQ1**, the quantitative document analysis revealed that students with different language proficiencies and technological competences displayed diversified developmental trajectories of feedback seeking abilities. First, students with higher language proficiency (i.e., Toby and Isabella) tended to create content-rich prompts in relatively fewer rounds of interaction with AI system, while learner with moderate language proficiency (i.e., Lisa) resorted to an opposite path. The differences in the quantity and length of prompts could be partially explained by the claim that the utilization and processing of feedback depended on the learners' language skills (Koltovskaia, 2020). Such phenomenon could also be attributed to the interrelation between L2 language mindset and FSB (Yao and Zhu, 2022). Second, judging from the time consumption of the participants, Toby and Isabella (representing learners with higher language proficiency) were relatively slow-paced in action. This echoed with the observation of Ranalli (2018) that AWCF would affect mental-effort expenditure on learners with different L2 levels. However, contrary to the claim that AWCF would pose extra cognitive overload to low-proficiency learners (Bitchener and Ferris, 2012), the current study revealed that it was high-proficiency students (i.e., Toby and Isabella) who spent more time and efforts analyzing, evaluating and making use the AWCF. Such contradiction could be attributed to the iterative and interaction mechanism of ChatGPT, in which low-proficiency learners could directly ask ChatGPT to regenerate (as evidenced in Lisa's overuse of *feedback regeneration* in **Figure 2**) while high-proficiency learners

(as evidenced in the dramatic decrease in feedback regeneration of Toby and Isabella) would stop, think, and write prompts with additionally backgrounds and specific demands (as evidenced by Toby and Isabella's prompt writing habits). Additionally, the variances in time consumption testified the existence of *inner feedback*, a natural process of processing and comparison when the learners was exposed to feedback (Nicol, 2021).

Internally, the weekly ChatGPT prompt patterns reflected the strategical aspect of feedback seeking ability development. First, students with different language proficiencies kept abreast in exerting feedback elicitation strategies in the interaction with ChatGPT. This showcased students' progress in understanding the conversational and iterative mechanism of ChatGPT. With the progression in the project, students gradually reduced writing minimalistic prompts, instead, additional information were included to enrich and clarify the instructions prepared for ChatGPT. This finding echoed the developmental characteristics of students using ChatGPT for text generation in an L2 classroom (Yan, 2023). Second, diversity to use feedback refinement strategies surfaced from the quantitative document analysis. Students with higher language proficiencies (i.e., Toby and Isabella) steadily shifted from the style of short and frequent interaction with the AI system to a more deliberated and information-rich manner. Conversely, Lisa, representing students with moderate language proficiency but higher command of technologies, consistently carried out a straightforward way to interact with the AI system. To some extent, the finding was in tandem with the argument of Havranek and Cesnik's (2001) that proficient learners could make better use of feedback. As feedback refinement strategies functioned on a higher order, such disparities could also be attributed to the variance in metacognitive regulatory skills. In this sense, such phenomenon could be explained by the positive correlation between metacognition and FSB (Xu, 2021).

In answering **RQ2**, the synthetization of participants' perceptions pinpointed the decisive factors for fostering feedback seeking abilities in a ChatGPT-enhanced environment. First, the consensus of participants emphasized learners' proactive involvement. Such belief was in tandem with the view that contemporary feedback was a process of self-generated meaning making (Jensen *et al.*, 2023). Based on such views, we infer that the significance of human efforts, in the era of AI-enhanced learning, should be uplifted rather than diminished. Second, the participants stressed the significance of metacognitive regulation. In the study, participants effectively combined the internal feedback processing and external feedback refinement to make better use of ChatGPT's feedback. The observation contradicted traditional views that feedback seeking encompassed limited metacognitive regulation other than the comparison of value and cost of potential feedback information (Ashford *et al.*, 2003). ChatGPT called for higher level of metacognition because students were allowed to state their affective and cognitive judgments. Such contradiction could be attributed to the *new feedback paradigm*, in which students' proactivity and agency were held pivotal (Jensen *et al.*, 2023). From another perspective, such phenomenon testified the reported effects of ChatGPT on the promotion of self-regulation skills (Wu, Lee, *et al.*, 2023).

Conclusions

The study investigated the development of feedback seeking abilities of L2 writers utilizing ChatGPT for corrective feedback. Furthermore, the research identified the key factors deciding students' development of feedback seeking abilities in the new learning environment.

The study delivered conceptual and pedagogical implications. First, a sociomaterial perspective towards feedback would trigger researchers to expand and

redevelop existing conceptualization in the field of instructional feedback (Gravett, 2022). The study revealed that ChatGPT, when actively and strategically utilized, offered a competitive solution for AWCF in L2 learning. Considering the growth in the application of AI, such domain-specific and in-depth investigations into the provision and utilization of feedback in innovative learning environments would be helpful for the sustained development of educational studies in the new era. Second, educators should consider the development of students' abilities to use state-of-the-art tools. The results showed that the quality and effectiveness of ChatGPT-generated feedback were closely related to students' feedback seeking abilities. However, to date, such abilities were understudied, let alone properly trained. There remained much to be done for educators, e.g., altering curriculum and assessment setups, (re)constructing a learning environment levering AI affordance and peer collaboration, providing training on AI skills, etc.

The study was not without limitations. First, the study took a relatively narrow stance towards student's learning behaviors in the new context. The study focused on human-AI interaction during the utilization of ChatGPT in L2 classrooms. Consequently, other key factors related to learning process and outcomes, e.g., achievement, engagement, motivation, acceptance, etc., were not examined. In the follow-up studies, educators and researchers could continue the exploration of the impact of ChatGPT on different aspects of pedagogy. Second, the reported findings were based on a three-participant case study and could not be easily generalizable. Considering the complexity and diversity of L2 learning, follow-up studies are encouraged to examine students' feedback seeking with a larger sample size. Third, the study intentionally adopted a student-centered perspective in research design. As a result, the effects of ChatGPT on teacher's roles and instructional approaches were not examined. In subsequent studies, researchers could posit the research settings in teacher-fronted L2 classrooms, where the

interaction between students and AI systems would be mediated by the teacher-student and peer interactions. Finally, the study is restricted by its duration. With a relatively short time span, the study delivered exploratory insights necessary to understand students' feedback seeking from GAI tools used in learning. However, due to the limited amount of collected data, the study did not apply analytical methods such as inferential statistical analysis, which would offer more evidence regarding the developmental trajectories of students' feedback seeking abilities. In subsequent studies, researchers are encouraged to conduct longitudinal studies for the deepened understanding of feedback seeking with AI-enhanced learning tools.

Acknowledgement

The author would like to thank the handling editor and all three anonymous reviewers who have offered constructive feedback for the paper's quality improvement.

Declaration of Interest

The author has no competing interests to declare.

Funding

This research was funded by ****, grant number ****.

References

Adams, G. (2019), "A narrative study of the experience of feedback on a professional doctorate: 'a kind of flowing conversation'", *Studies in Continuing Education*, Routledge, Vol. 41 No. 2, pp. 191–206, doi: 10.1080/0158037X.2018.1526782.

Ashford, S.J., Blatt, R. and VandeWalle, D. (2003), "Reflections on the Looking Glass: A Review of Research on Feedback-Seeking Behavior in Organizations", *Journal of Management*, Vol. 29 No. 6, pp. 773–799, doi: 10.1016/S0149-2063(03)00079-5.

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., *et al.* (2023), "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity", arXiv, doi: 10.48550/arXiv.2302.04023.

Barrot, J.S. (2021), "Using automated written corrective feedback in the writing classrooms: effects on L2 writing accuracy", *Computer Assisted Language Learning*, Routledge, pp. 1–24, doi: 10.1080/09588221.2021.1936071.

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

Birt, L., Scott, S., Cavers, D., Campbell, C. and Walter, F. (2016), "Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation?", *Qualitative Health Research*, SAGE Publications Inc, Vol. 26 No. 13, pp. 1802–1811, doi: 10.1177/1049732316654870.

Bitchener, J. and Ferris, D. (2012), *Written Corrective Feedback in Second Language Acquisition and Writing*, Routledge, New York ; London.

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Hodder Arnold, United Kingdom, Vol. 3 No. 2, pp. 77–101, doi: 10.1191/1478088706qp063oa.

Creswell, J.W. and Plano Clark, V.L. (2018), *Designing and Conducting Mixed Methods Research*, Third Edition., SAGE, Los Angeles.

Duff, P. (2010), *Case Study Research in Applied Linguistics*, Transferred to digital printing., Routledge, New York.

Fang, T., Yang, S., Lan, K., Wong, D.F., Hu, J., Chao, L.S. and Zhang, Y. (2023), "Is ChatGPT a Highly Fluent Grammatical Error Correction System? A Comprehensive Evaluation", arXiv, 4 April, doi: 10.48550/arXiv.2304.01746.

Gravett, K. (2022), "Feedback literacies as sociomaterial practice", *Critical Studies in Education*, Routledge, Vol. 63 No. 2, pp. 261–274, doi: 10.1080/17508487.2020.1747099.

Hattie, J. and Timperley, H. (2007), "The Power of Feedback", *Review of Educational Research*, American Educational Research Association, Vol. 77 No. 1, pp. 81–112, doi: 10.3102/003465430298487.

Havranek, G. and Cesnik, H. (2001), "Factors affecting the success of corrective feedback":, *EUROSLA Yearbook*, John Benjamins Publishing Company, Vol. 1 No. 1, pp. 99–122, doi: 10.1075/eurosla.1.10hav.

Hyland, K. and Hyland, F. (2019), "Contexts and Issues in Feedback on L2 Writing", in Hyland, F. (Ed.), *Feedback in Second Language Writing: Contexts and Issues*, 2nd ed., Cambridge University Press, Cambridge, pp. 1–22, doi: 10.1017/9781108635547.003.

Jensen, L.X., Bearman, M. and Boud, D. (2023), "Feedback encounters: towards a framework for analysing and understanding feedback processes", *Assessment & Evaluation in Higher Education*, Routledge, Vol. 48 No. 1, pp. 121–134, doi: 10.1080/02602938.2022.2059446.

Kaptelinin, V. and Nardi, B. (2018), "Activity Theory as a Framework for Human-Technology Interaction Research", *Mind, Culture, and Activity*, Routledge, Vol. 25 No. 1, pp. 3–5, doi: 10.1080/10749039.2017.1393089.

Kessler, M. (2023), "Written corrective feedback in an online community: A typology of English language learners' requests and interlocutors' responses", *Computers and Composition*, Vol. 67, p. 102752, doi: 10.1016/j.compcom.2023.102752.

Koltovskaia, S. (2020), "Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study", *Assessing Writing*, Vol. 44, p. 100450, doi: 10.1016/j.asw.2020.100450.

Leenknecht, M., Hompus, P. and Van Der Schaaf, M. (2019), "Feedback seeking behaviour in higher education: the association with students' goal orientation and deep learning approach", *Assessment & Evaluation in Higher Education*, Routledge, Vol. 44 No. 7, pp. 1069–1078, doi: 10.1080/02602938.2019.1571161.

Long, M. (1996), "The role of the linguistic environment in second language acquisition", in Ritchie, W. and Bhatia, T. (Eds.), *Handbook of Second Language Acquisition*, Academic Press, San Diego, CA, pp. 413–468.

Malecka, B., Boud, D. and Carless, D. (2022), "Eliciting, processing and enacting feedback: mechanisms for embedding student feedback literacy within the curriculum", *Teaching in Higher Education*, Routledge, Vol. 27 No. 7, pp. 908–922, doi: 10.1080/13562517.2020.1754784.

Mizumoto, A. and Eguchi, M. (2023), "Exploring the potential of using an AI language model for automated essay scoring", *Research Methods in Applied Linguistics*, Vol. 2 No. 2, p. 100050, doi: 10.1016/j.rmal.2023.100050.

Ngo, T.T.-N., Chen, H.H.-J. and Lai, K.K.-W. (2022), "The effectiveness of automated writing evaluation in EFL/ESL writing: a three-level meta-analysis", *Interactive Learning Environments*, Routledge, pp. 1–18, doi: 10.1080/10494820.2022.2096642.

Nicol, D. (2021), "The power of internal feedback: exploiting natural comparison processes", *Assessment & Evaluation in Higher Education*, Routledge, Vol. 46 No. 5, pp. 756–778, doi: 10.1080/02602938.2020.1823314.

Papi, M., Rios, A., Pelt, H. and Ozdemir, E. (2019), "Feedback-Seeking Behavior in Language Learning: Basic Components and Motivational Antecedents", *The Modern Language Journal*, Vol. 103 No. 1, pp. 205–226, doi: 10.1111/modl.12538.

Ranalli, J. (2018), "Automated written corrective feedback: how well can students make use of it?", *Computer Assisted Language Learning*, Routledge, Vol. 31 No. 7, pp. 653–674, doi: 10.1080/09588221.2018.1428994.

Stake, R.E. (1995), *The Art of Case Study Research*, Sage Publications, Thousand Oaks.

Stake, R.E. (2005), *Multiple Case Study Analysis*, Illustrated edition., The Guilford Press, New York.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., *et al.* (2023), "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT", arXiv, 21 February, doi: 10.48550/arXiv.2302.11382.

Wiliam, D. (2018), "Feedback: At the Heart of – But Definitely Not All of – Formative Assessment", in Lipnevich, A.A. and Smith, J.K. (Eds.), *The Cambridge Handbook of Instructional Feedback*, Cambridge University Press, Cambridge, pp. 3–28, doi: 10.1017/9781316832134.003.

Winstone, N.E., Hepper, E.G. and Nash, R.A. (2021), "Individual differences in self-reported use of assessment feedback: the mediating role of feedback beliefs", *Educational Psychology*, Routledge, Vol. 41 No. 7, pp. 844–862, doi: 10.1080/01443410.2019.1693510.

Wu, H., Wang, W., Wan, Y., Jiao, W. and Lyu, M. (2023), "ChatGPT or Grammarly? Evaluating ChatGPT on Grammatical Error Correction Benchmark", arXiv, 14 March, doi: 10.48550/arXiv.2303.13648.

Wu, T.-T., Lee, H.-Y., Li, P.-H., Huang, C.-N. and Huang, Y.-M. (2023), "Promoting Self-Regulation Progress and Knowledge Construction in Blended Learning via ChatGPT-Based Learning Aid", *Journal of Educational Computing Research*, SAGE Publications Inc, p. 07356331231191125, doi: 10.1177/07356331231191125.

Xu, J. (2021), "Chinese University Students' L2 Writing Feedback Orientation and Self-Regulated Learning Writing Strategies in Online Teaching During COVID-19", *The Asia-Pacific Education Researcher*, Vol. 30 No. 6, pp. 563–574, doi: 10.1007/s40299-021-00586-6.

Xu, J. and Wang, Y. (2023), "Individual differences in L2 writing feedback-seeking behaviors: The predictive roles of various motivational constructs", *Assessing Writing*, Vol. 56, p. 100698, doi: 10.1016/j.asw.2023.100698.

Yan, D. (2023), "Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation", *Education and Information Technologies*, Vol. 28 No. 11, pp. 13943–13967, doi: 10.1007/s10639-023-11742-4.

Yao, Y. and Zhu, X. (2022), "Modeling chinese high school EFL writers' language mindsets, goal orientations, and feedback-seeking behavior: the interaction effect between language mindsets and writing competence", *Reading and Writing*, doi: 10.1007/s11145-022-10339-6.

Yin, R.K. (2013), *Case Study Research: Design and Methods*, Fifth edition., SAGE Publications, Inc, Los Angeles.

Zhang, S. (2021), "Review of automated writing evaluation systems", *Journal of China Computer-Assisted Language Learning*, De Gruyter, Vol. 1 No. 1, pp. 170–176, doi: 10.1515/jccall-2021-2007.

Zhang, Z. and Hyland, K. (2023), "Student engagement with peer feedback in l2 writing: Insights from reflective journaling and revising practices", *Assessing Writing*, Vol. 58, p. 100784, doi: 10.1016/j.asw.2023.100784.