Quest Journals
Journal of Research in Humanities and Social Science
Volume 14 ~ Issue 5 (May 2026) pp: 42-48
ISSN(Online):2321-9467
www.questjournals.org

Research Paper

Making use of DeepSeek-generated feedback on EFL writing

Haitao Chen
School of Foreign Studies, Guangdong University of Finance and Economics
Corresponding Author: Haitao Chen

ABSTRACT
This study aimed to examine the effectiveness of DeepSeek-generated feedback on EFL writing based on a particularly devised prompt. Data comprised feedback, revisions, essay scores, perceptions and interviews. Analyses of DeepSeek-generated feedback revealed that DeepSeek was able to generate a large quantity of valid feedback concerning the language, content and organization of students' writing, but it excelled in language and organization feedback. Besides, the students utilized a vast majority of DeepSeek-generated feedback, particularly related to language and organization. Moreover, the students' revised drafts were significantly improved in language, content, organization and as a whole. Finally, they held a highly positive view of DeepSeek-generated feedback and were willing to use it for their future writing. Analyses of the focal participants' interviews revealed their positive attitudes towards DeepSeek-generated feedback with reservations. The findings of this study indicated that with a well-designed prompt, DeepSeek-like GenAI tools can generate effective feedback on EFL writing.

KEYWORDS: *DeepSeek-generated feedback, effective prompts, revisions, improvement, perceptions*

*Received 02 May., 2026; **R**evised 10 May., 2026; **A**ccepted 12 May., 2026 © The author(s) 2026.*
P**ublished with open access at **www.questjournals.org

I. INTRODUCTION

Since the release of ChatGPT in 2022, generative artificial intelligence (GenAI) tools have been gradually integrated into nearly all fields including English language teaching (Moorhouse 2024). As GenAI offers abundant affordances like quick delivery and personalized responses, it is expected to work as a dependable writing assistant tool for L2 student writers (Barrot 2023).

In the literature, a growing number of studies have been conducted to examine the effectiveness of GenAI-generated feedback on writing. Research has revealed that GenAI excels in dealing with grammatical errors and sentence problems, but is weak in providing feedback on higher-order issues. For example, Algaraady and Mahyoob (2023) revealed that the errors identified by ChatGPT were largely related to surface-level issues like tenses, punctuation and spelling.

GenAI tools are claimed to be able to generate personalized feedback (Barrot ibid.). However, research has revealed that GenAI was limited in its ability to provide feedback catering to individual differences. In Evmenova et al. (2024), ChatGPT provided the students in grades 3-7 with feedback on their English essays partly in line with teacher-specified areas of need and instructional decisions, but could not provide feedback on students' essays according to student characteristics like grade levels.

Research has also examined GenAI-generated feedback in terms of uptake and perceptions. Regarding students' uptake of ChatGPT-generated feedback, nearly 33% was not attempted (Zou et al. 2025), indicating a low uptake rate of the feedback. Concerning students' perceptions, they held positive perceptions of the contributions of GenAI-generated feedback to their EFL writing (Alsofyani and Barzanji 2025; Zou et al. 2025). Notably, only half of the students considered ChatGPT-generated feedback on L2 Spanish writing effective (Barrios-Beltran 2025).

The findings of the above-reviewed studies contribute to our understanding of GenAI-generated feedback on students' L2 writing, but the inconsistency in the findings warrants more research. As summarized in Table 1, the prompts of the above-reviewed studies differ in their requirements for GenAI tools to generate feedback. As GenAI tools generated feedback in line with the prompts received, the inconsistency in the findings relates largely

to the prompts for GenAI tools. Besides, few of the above-reviewed studies were targeted at the effectiveness of GenAI-generated feedback on L2 writing. Thus, it is worthwhile to investigate the effectiveness of GenAI-generated feedback on L2 writing based on a well-devised prompt. Such investigations can help develop a better understanding of how GenAI-generated feedback can benefit its receivers.

Table 1 Comparison of the prompts used in previous studies

Studies	Prompt
Alsofyani and Barzanji (2025)	Researchers designed tailored, detailed prompts for students aligned with the writing rubric
Evmenova et al. (2024)	Four prompts of varying specificity
Zou et al. (2025)	A general prompt
Chen et al. (2024)	Two prompts: one general and the other more specific
Guo and Wang (2024)	A general prompt comprising three sections
Barrios-Beltran (2025)	A detailed prompt with calibration sample essays

The effectiveness of GenAI-generated feedback meshes well with the Vygotskyan (Vygotsky 1978) concepts of zone of proximal development (ZPD) and scaffolding. ZPD is the distance between what a learner can solve a problem independently and what he/ she can work out a problem only with the help of more capable others (Vygotsky ibid.). Often used in conjunction with ZPD is the concept of scaffolding, which is described as "those supportive behaviors by which an expert can help a novice learner achieve higher levels of regulation" (de Guerrero and Villamil 2000: 51). For GenAI-generated feedback to be more effective, it is essential for the feedback to align with L2 students' ZPDs (Nassaji and Swain 2000). Those pieces of feedback out of their ZPDs will be perceived as useless and abandoned (Price et al. 2010).

In response to the gaps identified, in light of an easier access to the GenAI tool of DeepSeek in Chinese educational context, in line with the measures of the effectiveness of peer feedback in Hu and Lam (2010) (i.e. proportion of valid feedback, proportion of valid feedback taken up in revision, improvement in the second drafts and student perceptions of feedback), and informed by the sociocultural theory, four research questions were formulated for this study.

1. To what extent is DeepSeek-generated feedback valid?
2. How is the uptake of DeepSeek-generated feedback?
3. How does students' writing change after they make revisions based on DeepSeek-generated feedback?
4. How do the students perceive DeepSeek-generated feedback?

II. METHODS

2.1 Participants

A total of 29 participants were recruited from the English-major freshmen enrolled in an English writing course. As presented in Table 2, four students were selected to join semi-structured interviews. According to their scores for the *Oxford Quick Placement Test* (2001) they took before the study, their English proficiency was at intermediate levels. As the students had rarely consulted GenAI-generated feedback on their EFL writing and had an easier access to the GenAI tool of DeepSeek in China, their teacher provided them with a brief introduction to DeepSeek use in generating feedback on EFL writing. All students were informed of research purposes and procedures and written consent was attained from them before the study.

Table 2 Demographic information of the focal participants

Pseudonym	Gender	Age	Proficiency	Perceptions
Amy	Female	18	HP	Higher
Judy	Female	18	HP	Lower
Cherry	Female	18	LP	Higher
Kevin	Male	19	LP	Lower

2.2 Instruments

To gauge students' perceptions of DeepSeek-generated feedback, a 6-point Likert scale (see Appendix A) with answer choices anchored in 1 (strongly disagree) and 6 (strongly agree) was administered. The scale items were developed after a comprehensive review of relevant literature on feedback use in L2 writing (e.g. Escalante et al. 2023; Zou et al. 2025), and they were related to students' perceptions of the clarity, interpretability, usefulness of DeepSeek-generated feedback, and their willingness to use such feedback. Besides, semi-structured interviews (see Appendix B for an interview guide) were conducted individually with the focal participants to collect more in-depth information for their uptake and perceptions of DeepSeek-generated feedback.

Table 3 The prompt for DeepSeek to generate feedback on students' EFL writing

Aspect	Prompt requirements
Student information	Students are English major freshmen with English proficiency ranging from intermediate to upper intermediate.
Language	1. Provide feedback on vocabulary, grammar and mechanical issues;
Content	2. Provide feedback on the relevance of discussions in relation to major and minor propositions; 3. Provide feedback on the adequacy of discussions in relation to major and minor propositions;
Organization	4. Provide feedback on the organization of the major proposition in the whole essay; 5. Provide feedback on the organization of the minor propositions in corresponding paragraphs; 6. Provide feedback on the inter-sentence/ paragraph transitions; 7. Provide feedback on the overall coherence; 8. A summative feedback at the end is required;
Others	9. Feedback languages are English and Chinese; 10. Feedback is required to be presented in brackets; 11. Feedback comprises identified problems and/ or solutions provided.

2.3 An effective prompt for GenAI to generate feedback

To construct an effective prompt for GenAI tools to generate effective feedback, some principles should be observed: try to be detailed, provide context information, use simple language, clarify steps and engage in an iterative refinement process (Moorhouse op.cit.). Based on the above-introduced principles for effective prompts, the author of this manuscript and two experienced EFL writing teachers devised three initial prompts. After going through iterative revisions, the prompt was finalized with their agreement and presented in Table 3.

To maximize the benefits that the students could gain from DeepSeek-generated feedback, they were required to enter the different sections of the prompt in the order of content/ organization feedback requirement + language feedback requirement. That is, they revised their drafts first based on DeepSeek-generated content and organization feedback, and then they revised the language of their drafts based on the language feedback that DeepSeek generated on the drafts with revisions in content and organization.

2.4 Feedback effectiveness

In this study, the effectiveness of *DeepSeek*-generated feedback was examined in line with the four measures in Hu and Lam (2010): proportion of valid feedback, proportion of valid feedback taken up in revision, improvement in the second drafts and student perceptions of feedback. Valid feedback referred to appropriate feedback necessitating follow-up revisions. Evaluative feedback (i.e. praises) was excluded from valid feedback, while summative feedback was retained, as evaluative feedback did not necessitate immediate or long-term follow-up revisions. Except student perceptions, the other three measures were computed by category (i.e. language, content, and organization feedback) and across the categories for the students as a whole. Language feedback addressed problems related to grammar, vocabulary or mechanics. Content feedback was concerned with the adequacy and relevance of major or minor propositions of students' essays. Organization feedback dealt with problems concerning the development and organization of major or minor propositions.

2.5 Data collection

The data used in this study comprised feedback, revisions, essay scores, perceptions and interviews, and they were collected at four stages. After the students finished drafting in class, they entered in DeepSeek the prompt given by their teacher to seek feedback on their drafts. Then, they revised the drafts based on DeepSeek-generated feedback in class and handed in their first drafts, feedback, and revised drafts (i.e. second drafts). After that, they filled in an online questionnaire survey administered to assess their perceptions of DeepSeek-generated feedback. Finally, semi-structured interviews were conducted individually with the four particularly selected focal participants to elicit in-depth information about their uptake and perceptions of DeepSeek-generated feedback. The interviews were conducted in their mother tongue (i.e. Chinese) and each interview lasted around 25 minutes.

2.6 Data analysis

To answer the research questions, the author of this manuscript and an experienced EFL writing teacher coded feedback and revisions, and the inter-coder agreement reached 94% and 91% for valid feedback and valid revisions, respectively. Then they marked first and second drafts in line with a 100-point marking scheme developed from Paulus' (1999) scoring rubric. The marking scheme comprised assessment of language (maximum = 50 points), content (maximum = 30 points) and organization (maximum = 20 points). The scores provided by the two raters were averaged as the final scores for language, content and organization, and the sum of the averaged scores for the three aspects constituted the overall scores of students' drafts. The two raters calibrated their marking first by scoring six drafts independently, and then they marked all the first and second drafts with an acceptable maximum difference of 5 points in the overall score.

For the first research question, the quantities of feedback and valid feedback were counted, and the

proportion of valid feedback was calculated with the formula of (quantity of valid feedback/quantity of feedback) × 100. For the second research question, the quantities of valid feedback and valid feedback taken up were counted, and the proportion of valid feedback taken up in revision was calculated with the formula of (quantity of valid feedback taken up/quantity of valid feedback) × 100. For the third research question, multiple paired samples *t*-tests were performed on the scores of students' first and second drafts in terms of language, content, organization and the three categories as a whole. For the fourth research question, students' overall perceptions were calculated by averaging the scores for all scale items. Then, the scores of their perceptions were averaged for each item related to clarity, interpretability, usefulness of language feedback, usefulness of content feedback, usefulness of organization feedback, and willingness to use such feedback.

To analyse the semi-structured interviews, content analysis was conducted. The interviews were first transcribed verbatim and reviewed by the author multiple times. The author came up with initial codes and categorized them into three categories of feedback effectiveness: language-, content- and organization-related. After iterative checking, the codes and categories were finalized as the coding scheme. A second coder was first familiarized with the coding scheme and then coded the interviews independently. The intercoder agreement reached 90%. Any disagreement in the coding was solved with negotiation. The first coder further analysed the coding results and identified the reasons for students' uptake and perceptions of DeepSeek-generated language, content and organization feedback.

III. RESULTS

DeepSeek generated 621 pieces of feedback in total for all 29 writing drafts. The feedback was distributed across three aspects: language, content and organization. Of the feedback, 558 were deemed valid. As presented in Table 4, DeepSeek generated large quantities of valid feedback, and the proportions of valid feedback ranged from 71% to 96% for content, organization and language feedback.

Besides, the students used a large quantity of the valid feedback with the overall proportion of valid feedback taken up in revision reaching 87% and the proportions of valid organization, content and language feedback reaching 92%, 71% and 86%, respectively.

Table 4 Proportions of valid feedback and uptake rates by type and in aggregate

Measure	Language	Content	Organization	Total
Quantity of feedback	347	142	132	621
Quantity of valid feedback	332	101	125	558
% of valid feedback	95.68%	71.13%	94.70%	89.86%
Quantity of valid feedback taken up	304	72	107	483
% of valid feedback taken up	91.57%	71.29%	85.60%	86.56%

The scores for students' first and second drafts were compared. As summarized in Table 5, the average scores for the language, content and organization of their writing increased from first to second drafts. The results of paired samples *t*-tests revealed that the students made significant improvement not only in their overall writing scores but also in the scores for the language, content and organization of their writing, with small to large effect sizes.

Table 5 Results of paired-samples *t*-tests comparing first and second drafts

Aspect	1st drafts		2nd drafts		*df*	*t*	*p*	Cohen's *d*
	Mean	*SD*	*Mean*	*SD*			*(2-tailed)*	
Language	30.55	5.42	32.03	4.69	28	-6.72	.000	1.25
Content	18.69	2.35	19.04	2.09	28	-2.47	.020	.46
Organization	12.41	2.38	14.22	2.28	28	-8.33	.000	1.55
Overall	61.65	8.59	65.29	7.44	28	-9.53	.000	1.77

Regarding students' perceptions of DeepSeek-generated feedback, the average score for their overall perceptions reached as high as 5.41. As presented in Table 6, the scores for each scale item exceeded 4.97. Specifically, the average scores for their perceptions of feedback usefulness descended from language to organization to content feedback. The average score for their willingness to use DeepSeek-generated feedback reached 5.72.

Table 6 Students' perceptions of DeepSeek-generated feedback

Item	*DeepSeek*-generated	
	M	*SD*
1. The clarity of feedback	5.52	.51
2. The interpretability of feedback	5.41	.63
3. The usefulness of language feedback	5.59	.50

4. The usefulness of content feedback	4.97	.78
5. The usefulness of organization feedback	5.28	.70
6. Willingness to continue using feedback	5.72	.45

Analyses of the focal participants' interviews revealed their positive attitudes towards DeepSeek-generated feedback with some reservations. First, they all admitted the well-targetedness of DeepSeek-generated three types of feedback as the feedback pointed out their weaknesses. For example, Judy and Kevin appreciated the language feedback which helped them know why some words and sentence structures in their writing were erroneous and how to correct such errors. Second, they appreciated the guidance provided for their future writing improvement. The guidance was included in some evaluative and summative feedback. As they said, the guidance pointed out "the directions that I should make more efforts in".

Third, they held more reservations about content feedback than language and organization feedback. Regarding language feedback, they all deemed a few suggested academic words beyond their capabilities. With respect to organization feedback, some feedback lacked specific suggestions. For example, a piece of feedback in Kevin's draft read, "This minor proposition was not well-supported and you were advised to rewrite it". Concerning content feedback, they considered some feedback too general. Judy and Cherry did not act upon some content feedback as those pieces of feedback were "too general and difficult to cope with". They also deemed some feedback inappropriate due to the change of their original information. Amy and Cherry said, the logic in some content feedback could stand, but those pieces of feedback changed "What I wanted to express".

Lastly, they expected guidance from summative feedback with more specific and tailored suggestions. They needed more detailed suggestions on how to improve their writing rather than "simple and empty suggestions that just remind me of heeding word choice, grammar, coherence and paragraph development" (Cherry, interview). Judy, Cherry and Kevin opined that general suggestions were not suitable for them but may fit students of advanced English writing proficiency.

IV. DISCUSSION

Adopting a mixed-methods approach, this study examined the effectiveness of GenAI-generated feedback on EFL writing based on a particularly devised prompt. Most of the DeepSeek-generated feedback was revealed to be valid. Specifically, the proportions of valid language and organization feedback reached 96% and 95%, respectively, providing reassuring evidence that DeepSeek was expert at locating rule-based problems and providing corresponding suggestions. Meanwhile, the proportion of valid content, lower than the other two types of feedback, reached 71%, indicating that GenAI was weaker but still had a role to play in dealing with higher-order issues like recognizing content problems.

Besides, the participants acted upon most DeepSeek-generated feedback and improved their writing drafts. The proportions of feedback taken up in revisions reached 87%, higher than the proportions reported in Zou et al. (2025). The revisions based on the feedback contributed to the students' writing improvement. Specifically, their second drafts were significantly improved in the language, organization and content and as a whole from their first drafts. These findings indicated that the DeepSeek-generated feedback largely met the students' needs and could be managed by them as it aligned with their ZPDs.

Moreover, they were highly satisfied with and were strongly willing to use DeepSeek-generated feedback. Similar to the findings in earlier studies (Alsofyani & Barzanji op.cit.; Zou et al. op.cit.), they appreciated the usefulness of language, organization and content feedback. Their satisfaction, as explained in the interviews, pertained to the overall well-targetedness of language, organization and content feedback, although they were also confronted with difficulties when engaging in the three types of feedback. These findings do not support the findings reported in Evmenova et al. (2024). One possible explanation for this discrepancy lies in the difficulties of providing feedback for receivers. The receivers in their study were students in grades 3-7 with different learning disabilities, while the receivers in this study were university students. The well-targetedness of language, content and organization feedback fitted well with their ZPDs, made them recognize the usefulness of DeepSeek-generated feedback and strengthened their willingness to use such feedback on their future writing tasks.

Finally, their successful uptake proportions and average perception scores of the feedback descended from language to organization to content feedback. As explained in the interviews, this trend can be contributed to the ease of using language and organization feedback and the more difficulties in adopting content feedback. Although the three types of feedback were useful, the more difficulties made content feedback less likely to constitute scaffolding, hence perceived as less useful and less acted upon.

Overall, the findings of this study provided evidence of the effectiveness of DeepSeek-generated feedback on students' EFL writing based on a particularly devised prompt. According to the interview findings, the participants were satisfied with DeepSeek-generated feedback because the feedback was well-targeted at their weaknesses and provided guidance for their future development, indicating that the scaffolded feedback generally

meshed well with their ZPDs. Put differently, the feedback not only helped address their language, content and organization problems, but also pointed to the directions that they should make more efforts in.

The findings of this study have some pedagogical implications. First, devising a tailored prompt for GenAI tools is of paramount importance. In this study, GenAI generated effective feedback based on a validated prompt specifying the writing context, the feedback needed and students' proficiency. Those L2 writing teachers wanting their students to harness GenAI-generated feedback should prepare a well-devised prompt with specific requirements and tailored to their students' proficiency. Second, students should make differentiated use of GenAI-generated language, content and organization feedback. As revealed in this study, GenAI-generated three types of feedback was generally effective with some difficulties, so students should be reminded of using the feedback with due caution. Third, teachers are encouraged to provide timely help for their students if GenAI is used in their L2 writing classes. As students were not able to deal with some suggestions in the feedback like some academic words, the general scopes of some content and organization feedback, teachers are advised to collect problematic feedback from their students and provide help when necessary.

V. CONCLUSION

This study examined the effectiveness of GenAI-generated feedback on EFL writing based on a particularly devised prompt. It was revealed that the feedback was valid with much feedback utilized, and that the students were satisfied with the feedback and were willing to use such feedback on future writing. These findings substantiated the effectiveness of GenAI-generated feedback on EFL writing.

Admittedly, this study has several limitations. First, the students in this study used GenAI-generated feedback only for one writing task. Such a one-shot design can not capture changes in students' uptake and perceptions of the feedback. Meanwhile, it does not allow the prompt to be further refined to provide more tailored feedback for the students. Second, only four students were involved in the semi-structured interviews. This small number of participants could not represent all the students involved in this study, hence impacting on the strength of the generalizability of the interview findings.

Funding Projects:
This research was supported by Guangdong Provincial Education Science Planning Project (Departmental Level) – Special Project for Higher Education in 2023 (2023GXJK293).

Appendix A. Students' Perceptions of DeepSeek-generated Feedback

1. DeepSeek-generated feedback was clear.
2. DeepSeek-generated feedback was easy to understand.
3. DeepSeek-generated language feedback was useful.
4. DeepSeek-generated content feedback was useful.
5. DeepSeek-generated organization feedback was useful.
6. I will use DeepSeek-generated feedback for future writing.

Appendix B. A Semi-Structured Interview Guide

1. How do you think of DeepSeek-generated language feedback?
2. How do you think of DeepSeek-generated content feedback?
3. How do you think of DeepSeek-generated organization feedback?
4. How do you think of DeepSeek-generated summative feedback?

REFERENCES

[1]. Algaraady, J. and M. Mahyoob. 2023. 'ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners'. *Arab World English Journals, Special Issue on CALL* 9: 3–17.
[2]. Alsofyani, A. H., and A. M. Barzanji. 2025. 'The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study'. *Journal of Educational Computing Research* 63/2: 431-463.
[3]. Barrios-Beltran, D. 2025. 'Exploring the efficacy of ChatGPT-4 feedback in second language Spanish writing'. *System* 133: 103771.
[4]. Barrot, J. S. 2023. 'Using ChatGPT for Second Language Writing: Pitfalls and Potentials.' *Assessing Writing* 57:100745.
[5]. de Guerrero, M. C. M., and O. S. Villamil. 2000. 'Activating the ZPD: Mutual scaffolding in L2 peer revision'. *Modern Language Journal* 84/1: 51–68.
[6]. Escalante, J., A. Pack, and A. Barrett. 2023. 'AI-generated feedback on writing: Insights into efficacy and ENL student preference'. *International Journal of Educational Technology in Higher Education* 20/1: 1-20.
[7]. Evmenova, A. S., K. Regan, R. Mergen, and R. Hrisseh. 2024. 'Improving writing feedback for struggling writers: Generative AI to the rescue?'. *TechTrends* 68/4: 790-802.
[8]. Hu, G., and S. T. E. Lam. 2010. 'Issues of cultural appropriateness and pedagogical efficacy: Exploring peer review in a second language writing class'. *Instructional Science* 38/4: 371–394.
[9]. Nassaji, H., and M. Swain. 2000. 'A Vygotskian perspective on corrective feedback in L2: The effect of random versus negotiated help on the learning of English articles'. *Language awareness* 9/1: 34-51.
[10]. Moorhouse, B. L. 2024. 'Generative artificial intelligence and ELT'. *ELT Journal* 78/4: 378-392.
[11]. Oxford University. 2001. *Quick placement test (paper and pen user manual)*. Oxford: Oxford University Press.

[12]. Paulus, T. M. 1999. 'The effect of peer and teacher feedback on student writing'. *Journal of second language writing* 8/3: 265-289.
[13]. Price, M., K. Handley, J. Millar, and B. O'donovan. 2010. 'Feedback: all that effort, but what is the effect?'. *Assessment & Evaluation in Higher Education* 35/3: 277-289.
[14]. Vygotsky, L. S. 1978. *Mind in society: The development of higher mental process*. Cambridge: Harvard University Press.
[15]. Zou, S., K. Guo, J. Wang, and Y. Liu. 2025. 'Investigating students' uptake of teacher-and ChatGPT-generated feedback in EFL writing: A comparison study'. *Computer Assisted Language Learning*:1–30.