

The Interpreter and Translator Trainer

ISSN: 1750-399X (Print) 1757-0417 (Online) Journal homepage: www.tandfonline.com/journals/ritt20

Adapt or adopt? Examining the efficacy of ChatGPT in providing translation feedback

Yanfang Su, Simin Xu & Kanglong Liu

To cite this article: Yanfang Su, Simin Xu & Kanglong Liu (11 Aug 2025): Adapt or adopt? Examining the efficacy of ChatGPT in providing translation feedback, The Interpreter and Translator Trainer, DOI: 10.1080/1750399X.2025.2541486

To link to this article: https://doi.org/10.1080/1750399X.2025.2541486

Published online: 11 Aug 2025.

Submit your article to this journal

Article views: 52

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ritt20

THE INTERPRETER AND TRANSLATOR TRAINER
https://doi.org/10.1080/1750399X.2025.2541486

Routledge
Taylor & Francis Group

Check for updates

Adapt or adopt? Examining the efficacy of ChatGPT in providing translation feedback

Yanfang Su, Simin Xu and Kanglong Liu

Department of Language Science and Technology, The Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

Translation feedback is crucial for improving student translation quality and cultivating translation competence, yet it remains a labour-intensive and time-consuming task for teachers. This study explores the potential of ChatGPT, a generative artificial intelligence (GenAI) model, as an alternative or supplement to traditional teacher feedback in translation education by comparing the content and strategy of these two feedback sources, as well as students' adoption and perception of these sources. A mixed-methods crossover experimental design was employed to assess feedback from ChatGPT and teachers across four translation tasks. The findings indicated that ChatGPT tended to offer direct reference translations and implicit general feedback, focusing predominantly on lexical and morpho-syntactic issues. In contrast, teacher feedback was more indirect and elaborated. Students showed a higher rate of acceptance with teacher feedback than with ChatGPT feedback. Interviews revealed that students held positive views towards teacher feedback but expressed mixed attitudes towards ChatGPT feedback regarding its quality, reliability, comprehensibility, and affective connection. The findings contribute to the discourse on the transformative role of GenAI in translation education and offer insights for integrating ChatGPT in translation teaching practice.

ARTICLE HISTORY
Received 17 December 2024
Accepted 28 July 2025

KEYWORDS
Translation feedback; ChatGPT; translation education; student perception; translation experiment

1. Introduction

Translation feedback plays a pivotal role in improving the quality of student translations and fostering their translation competence (Behr 2015). However, teachers often find it labour-intensive and time-consuming to provide timely and sufficiently constructive translation feedback (Korol 2022). These challenges highlight the necessity of investigating alternative feedback methods. The recent advancement in generative artificial intelligence (GenAI) models, particularly ChatGPT, offer a promising avenue. With its advanced natural language comprehension and capability to support coherent conversational flows (Floridi 2023), ChatGPT can offer customised, real-time feedback, assuming that suitable evaluation rubrics and precise instructions are incorporated into the prompts (Shadiev and Feng 2023). While researchers have investigated the potential of ChatGPT for feedback in the context of EFL writing (Guo and Wang 2024; Lu et al. 2024;

CONTACT Kanglong Liu kl.liu@polyu.edu.hk
© 2025 Informa UK Limited, trading as Taylor & Francis Group

Steiss et al. 2024), its application in translation remains relatively unexplored. Therefore, this study seeks to address these gaps in the literature by conducting a detailed examination and comparison of the content and strategies of ChatGPT and teacher feedback, as well as students' adoption and perceptions of the feedback. The research contributes to understanding how automated feedback may differ from and complement human feedback. The findings offer teachers empirical insights into the development of translation feedback mechanisms and help students critically engage with ChatGPT's strengths and limitations to build feedback literacy in the age of generative AI.

2. Related work

2.1. Feedback in translation education

Effective feedback improves learning by setting clear performance standards, providing learners with insights into their progress, fostering self-reflection, and enhancing interactions between teachers and students, thus creating a nurturing learning environment (Hattie and Timperley 2007). In translation education, providing feedback on student translations is a widely adopted practice to improve students' translation quality and foster translation competence (Behr 2015). Since translation is a fundamentally dialogical process, translation feedback creates a dialogical space, where students can adopt a reader's perspective to understand translation quality standards and critically assess their work through multiple rounds of receiving and responding to feedback (Washbourne 2014). This step-by-step revision process allows students to progressively refine their translations based on reader needs and expectations (ibid.).

Previous research has underscored the benefits of teacher feedback in translator training. For instance, Bruton (2007) found that explicit translation feedback could facilitate vocabulary learning with the help of a dictionary. Yu et al. (2020) noted that teachers' written corrective feedback positively influenced students' translation accuracy and recommended using diverse feedback strategies to maximise learning benefits. Neunzig and Tanqueiro (2005) discovered that personalised and immediate feedback from teachers could positively affect students' attitudes towards translation teaching. Despite the benefits, providing feedback can be time-consuming, mentally exhausting, and effort-draining for teachers, especially when the class size is large (Lee 2019). Such pressure might make it difficult for teachers to provide the diversified, customised, and timely feedback that students expect (Han and Lu 2023).

The challenges faced by teachers underscore the need for alternative automated approaches to providing translation feedback. For instance, some researchers have confirmed the efficacy of using a grammar checker in improving the grammatical accuracy of students' translation, such as WhiteSmoke (Tabrizi and Etemad 2021) and Grammarly (Sofyan and Nasution 2023). However, since translation is a complex process that extends beyond linguistic accuracy, grammar checkers may not fully meet the needs of translation students. Some other scholars have proposed using automated machine translation metrics, such as BLEU, NIST, and METEOR, as automated feedback for translation and interpreting students (Han and Lu 2023). Their analyses have revealed a significant correlation between the scores generated by these automated metrics and human evaluations. Still, automated metrics cannot fully replicate the accuracy or depth

of human assessments. They have failed to offer detailed, constructive feedback about students' strengths, weaknesses, and areas for improvement. Thus, there remains a clear need for more sophisticated automated feedback mechanisms capable of providing more comprehensive feedback that aligns with the multiple competencies required in translation practice.

2.2. The efficacy of ChatGPT in providing feedback

Recent advancements in GenAI models, particularly those in the GPT series, offer promising solutions to alleviate the burden on teachers in the teaching process. Researchers have explored the potential of ChatGPT for providing written feedback. For example, Su, Lin, and Lai (2023) proposed using ChatGPT to generate feedback during various stages of the argumentative writing process, such as outline preparation, content revision, and language polishing. However, they also noted that ChatGPT might produce nonsensical or overly generic feedback that required further clarification. Guo and Wang (2024) compared feedback from ChatGPT to that from teachers and found that ChatGPT produced a larger volume of feedback with a more balanced focus on different aspects of content. ChatGPT provided both direct feedback, which specifically outlined necessary revisions, and general feedback, which assessed the overall quality of the writing. In contrast, teachers typically offered more indirect suggestions. Their study also highlighted teachers' mixed opinions about the utility of ChatGPT in providing feedback. Lu et al. (2024) complemented Guo and Wang's (2024) study by incorporating students' perspectives. They revealed that the scores generated by ChatGPT aligned with those given by teachers, yet students still favoured teacher feedback. They argued that while ChatGPT's feedback was sometimes vague, inaccurate, and impersonal, its variety could complement teacher feedback by improving students' feedback literacy. Steiss et al. (2024) evaluated the quality of human and ChatGPT feedback on students' writing, considering factors such as adherence to criteria, clear direction, accuracy, essential features, and supportive tone. Human evaluators outperformed ChatGPT in all dimensions except adherence to criteria. However, they considered ChatGPT a valuable alternative that balances efficiency and quality.

Existing research suggests that while ChatGPT feedback has certain limitations, it demonstrates the potential to be integrated with traditional teacher feedback. However, despite encouraging findings in writing, the application of ChatGPT in translation teaching remains largely unexplored. An exception is the research by Cao and Zhong (2023). The authors found that while teacher feedback and self-feedback improved BLEU scores, ChatGPT feedback enhanced lexical complexity and referential cohesion, as measured by Coh-Metrics indices. However, their use of simple prompts may not fully utilise ChatGPT's capabilities, and their focus on linguistic features may not adequately capture translation quality. The characteristics of feedback provided by teachers, ChatGPT, and students, as well as how students apply these types of feedback in their revisions, remain to be investigated. Therefore, the current study seeks to address these gaps by conducting a detailed examination of ChatGPT and teacher feedback and exploring how students utilise and perceive different sources of feedback. Specifically, the following research questions will be addressed:

(1) How does feedback generated by ChatGPT differ from feedback provided by teachers?
(2) To what extent do students utilise and incorporate feedback from ChatGPT and teachers when revising their translations?
(3) What are students' perceptions of the utility of feedback from ChatGPT compared to teacher feedback?

3. Methods

3.1. Context and participants

The research was carried out within the context of the Master of Arts in Translation and Interpreting (MATI) programme at the Hong Kong Polytechnic University. The study was part of an AI-assisted teaching project that aimed to enhance student AI literacy. To facilitate the project, an AI-powered online translation platform was developed, featuring various functions such as machine translation, language polishing, feedback generation, and assignment submission (Figure 1). This study specifically utilised the feedback generation function, which employed the ChatGPT 3.5 model to provide feedback on student translations. When using this function, teachers had the option to either provide feedback that was automatically generated by ChatGPT or upload feedback that they had created themselves. The process of designing the prompt for ChatGPT feedback involved multiple rounds of testing and revision to ensure a balance between reliability and creativity in the generated feedback. In creating the ChatGPT prompt for producing translation feedback, we followed the guidelines proposed by Giray (2023) for effective prompts. These guidelines included defining ChatGPT's role, describing the task context, incorporating the input data, and specifying the output format. To ensure consistency in the evaluation process, both ChatGPT and human evaluators were given a standardised assessment rubric adapted from Hurtado Albir (2015). This rubric comprised a range of criteria spanning various categories, including accuracy, clarity, orthography, lexis, morphosyntax, coherence, cohesion, and adherence to genre conventions.

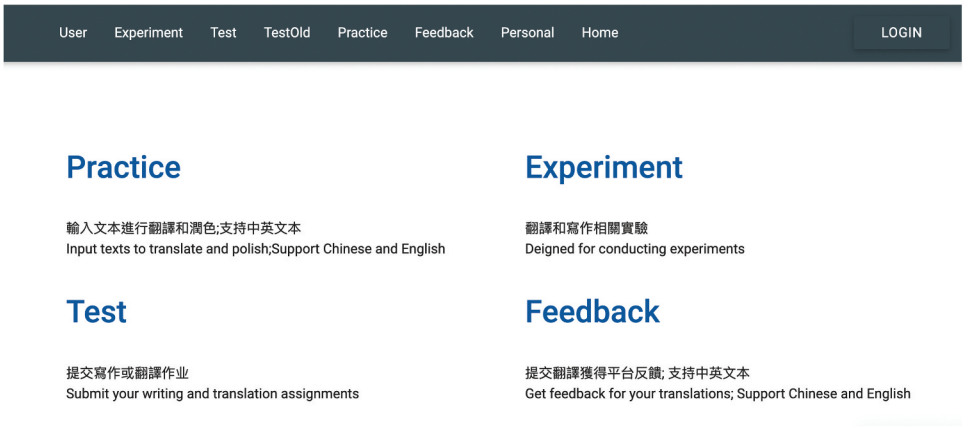

Figure 1. Screenshot of the AI-powered online translation system.

A total of 58 students from the one-year MATI programme voluntarily participated in the study after providing informed consent, representing approximately 60% of the cohort for that academic year. All participants were at the same stage in their studies, having completed the same core courses in Chinese-to-English and English-to-Chinese translation during the same timeframe. The participants were randomly divided into two groups labelled G1 and G2. A background survey confirmed that all participants were native Chinese speakers and learners of English as a Second Language (ESL), with an average IELTS score of 7.5. Their limited experience in using ChatGPT for translation feedback was a primary reason for their participation in the study. Prior to the experiment, each participant completed a Chinese-English translation placement assessment using materials from a simulated China Accreditation Test for Translators and Interpreters (CATTI). The results of the Mann-Whitney U test showed no significant difference between G1 (median = 77.45, SD = 3.39) and G2 (median = 77.72, SD = 4.09) in terms of their initial proficiency ($U = 407.5$, $p = 0.85$).

3.2. Research procedure

A mixed-methods crossover experimental design was utilised in this study, comprising four Chinese-to-English translation tasks. The translation direction was intentionally chosen based on two key considerations. First, as the participants were native Chinese speakers, additional support was necessary for producing English output. Second, ChatGPT 3.5 exhibits superior performance in generating natural-sounding English output compared to Chinese, likely due to the composition of its training data. To mitigate practice effects, translation tasks from various genres were included, including political news (Task 1), one business letter (Task 2), a travel narrative (Task 3), and one government report (Task 4). In order to ensure consistency and appropriateness of the task difficulty level, the framework for assessing translation difficulty (Sun 2015) was followed when selecting the translation texts. This process involved an initial selection of texts with comparable readability levels, as determined by the Chinese Readability Index Explorer (CRIE), which covers multiple lexical and syntactic indices. Then, two translation teachers and one research assistant reviewed the texts to identify and ensure consistency in potential translation challenges, thereby verifying their suitability for MATI students. Furthermore, an MATI student helper trial-translated all four tasks to ensure that each task could be completed within approximately one hour.

To ensure that students were well-prepared for the experiment, students attended two one-hour introductory sessions prior to the translation tasks. Each session helped students to learn ChatGPT basics, as well as the functionalities of the AI-powered translation platform. After completing these sessions, participants were given two weeks to complete and revise each translation task. During the first week, students had 70 minutes to complete their translations. In the subsequent week, they had an entire week to revise their translations based on the feedback they received, which was provided either by teachers or ChatGPT, depending on their group assignment. For the first two tasks, G1 received feedback from ChatGPT and G2 received teacher feedback. In the final two tasks, the feedback sources were switched, with G1 receiving teacher feedback and G2 receiving ChatGPT-generated feedback. The entire study lasted eight weeks.

Two translation teachers and one research assistant provided feedback on students' translations. In order to ensure the reliability and consistency of the feedback provided by the teacher, the three assessors established uniform criteria for providing feedback prior to the start of the experiment. Throughout the experiment, each assessor was responsible for delivering feedback on one-third of the student translations and cross-checking each other's feedback to maintain quality and consistency.

3.3. Data analysis

3.3.1. Comparing ChatGPT and human feedback

To address the first research question, we began the analysis by comparing ChatGPT-generated feedback with that provided by teachers across all four translation tasks. As ChatGPT's feedback typically includes a reference translation along with commentary, the researchers started by comparing the reference translations with the students' original translations to identify the revisions suggested by ChatGPT. Following this, the researchers segmented the feedback from both ChatGPT and human evaluators into discrete idea units, with each unit defined as a single remark addressing one specific aspect of the text (Link, Mehrzad, and Rahimi 2022).

After identifying the idea units in the feedback, the researchers categorised the feedback in terms of content and strategies. Feedback content refers to 'how well a task is being accomplished or performed' (Hattie and Timperley 2007, 91). The coding scheme for feedback content was adapted from previous research (Hurtado Albir and Pavani 2018; Li et al. 2023) and aligned with the assessment rubrics provided to both ChatGPT and human evaluators. Feedback strategies, on the other hand, were defined as 'the different degrees of correction and assignment of responsibility' for the translation (Washbourne 2014, 242). The coding scheme for feedback strategies was adapted from Ellis (2009) and Li et al. (2023). To ensure the reliability of the coding process, two authors initially analysed 16 pieces of translation (approximately 7% of the total data) as coders, achieving a consistency rate of 93.84%. Any discrepancies were resolved, and ambiguities were clarified through discussion. Following this, the two coders independently coded the remaining feedback and cross-checked each other's coding. The finalised coding schemes for feedback content and strategies are presented in Tables 1 and 2, respectively.

Following the coding process, the researchers counted the frequency of each category of feedback content and strategies provided by both ChatGPT and teachers. Descriptive statistics were then employed to illustrate the distributional patterns of the feedback. To confirm the potential statistical divergence between the feedback provided by ChatGPT and that provided by human evaluators, Chi-square Tests of Independence were conducted for both feedback content and strategies. Since the expected frequency within each cell of the contingency table should be at least five, the analysis of feedback strategies was limited to overarching categories rather than individual subcategories.

3.3.2. Examining student utilization of feedback

To address the second research question, we applied a coding scheme adapted from Li et al. (2023). Students' revision behaviours were quantified by identifying distinct

Table 1. Coding scheme for feedback content [adapted from Li et al. (2023); Hurtado Albir and Pavani (2018)].

Category	Codes	Description
Orthographic issues	Written	Comments on orthographic and typographic features
Lexis/wording	Le	Comments on accuracy and diversity of vocabulary
Morphosyntax	Morph	Comments on the use of verb tenses, modes, prepositions, and syntactic structures
Cohesion	Cohesion	Comments on the employment of cohesive devices and referential elements
Coherence	Coherence	Comments on how well the ideas are organized and clearly presented
Information accuracy and clarity	Info	Comments on the accuracy and clarity of the meanings conveyed
Genre conventions	Genre	Comments on the appropriateness of genre conventions and the translation purpose and target audience
Overall quality	Q	Comments on the overall quality of the translation

Table 2. Coding scheme for translation feedback strategies [adapted from Ellis (2009); Li et al. (2023)].

Category	Subcategory	Code	Description
Direct feedback	With explanation	Direct1	Correct the problem directly in the feedback with an explanation.
	Without explanation	Direct0	Correct the problem directly in the feedback without explanation.
Indirect feedback	Indication	Indirect1	Indicate the presence of an issue only.
	Indication + location	Indirect2	Indicate the presence of an issue and locate the issue.
	Indication + location + explanation	Indirect3	Indicate the presence of an issue, locate the issue, and include an explanation.
Reformulation	With explanation	Re1	Retranslate the entire text with an explanation.
	Without explanation	Re0	Retranslate the entire text without explanation.
Generic commentary		Gen	Provide a generic comment on the translation.

changes made to the original translation in response to feedback, with each modification targeting a specific translation problem counted as one revision unit. Two coders first coded 16 translations and achieved a consistency rate of 86.75%. Any differences were resolved, and uncertainties were clarified through discussion. Following this, the coders independently coded the remaining translation revisions (see Table 3 for the coding scheme). Descriptive statistics were used to illustrate students' uptake of feedback provided by either ChatGPT or human evaluators when revising their translations. Subsequently, Chi-square Tests of Independence were conducted to examine whether the utilisation of feedback varied significantly across different feedback providers, i.e., ChatGPT or human evaluators. Throughout the study, Python was used to generate descriptive plots and perform chi-square tests.

Table 3. Categories of revision [adapted from Li et al. (2023)].

Revision Act	Codes	Description
No change	NO	Students did not make revision attempts in response to the feedback.
Effective acceptance	EA	Students effectively accepted the feedback to make translation revisions.
Effective substitution	ES	Students successfully made their own revisions in response to the feedback.
Ineffective revision related to feedback	IR	Students made incorrect revisions to their translations in response to the feedback.
Effective self-initiated revision	ESR	Students recognized the need to make revisions and initiated the revision without reference to explicit feedback.
Ineffective self-initiated revision	ISR	Students initiated incorrect revisions to their translations without referring to the explicit feedback

3.3.3. Analyzing student perception of feedback

To address the third research question, semi-structured interviews were conducted with focal students. The purpose of the interview is to provide a 'thick description' and in-depth analysis of participants' lived experiences, thoughts, and feelings (Cohen, Manion, and Morrison 2011, 290) towards teacher and ChatGPT feedback, which were not directly observable. It can complement and explain the quantitative analysis of student revision behaviours (Patton 2014). Specifically, six students (Q, N, and L from G1; W, F, and X from G2) were purposefully selected for interviews after the experiment based on their active participation in all training sessions and translation tasks and willingness to be interviewed (cf. Patton 2014). This selection criterion aimed to ensure they could provide relatively objective feedback evaluations and detailed experience. During the interview, a stimulated-recall procedure was used to elicit their perceptions while they reviewed the feedback received and their translation revisions.

The interview transcripts were analysed using inductive thematic analysis to identify, analyse, and organise themes related to students' perceptions of the feedback, following the steps outlined by Nowell et al. (2017). To ensure the credibility of the thematic analysis and the comprehensiveness of the information obtained, both the first and second authors were involved in the interview and analysis process.

4. Results

4.1. Differences and similarities between ChatGPT and teacher feedback

4.1.1. Feedback content

A chi-square test of independence revealed a significant association between the feedback content and the feedback provider (GPT or human) in each of the four tasks (Task 1: χ^2 (7, 1149) = 120.06, $p < 0.001$; Task 2: χ^2 (7, 1260) = 146.01, $p < 0.001$; Task 3: χ^2 (7, 1194) = 39.40, $p < 0.001$; Task 4: χ^2 (7, 1248) = 128.95, $p < 0.001$), as well as the combined data across all four tasks (χ^2 (7, 4851) = 311.63, $p < 0.001$). Tables 4 and 5 present detailed descriptive statistics of the feedback content provided by ChatGPT and human evaluators, respectively.

Tables 4 and 5 reveal both similarities and differences in the content of feedback provided by ChatGPT and teachers. For both, the top four priorities were issues related to genre conventions, information clarity and accuracy, lexical choice, and morphosyntax. Comparatively, less emphasis was placed on the coherence and cohesion of translations, orthographic issues, and the overall quality of the translations.

Table 4. Feedback content provided by ChatGPT*.

Feedback Type	Task 1	Task 2	Task 3	Task 4	Total
Coherence	24 (3.09%)	29 (3.38%)	30 (4.43%)	24 (3.32%)	107 (3.53%)
Cohesion	49 (6.31%)	68 (7.92%)	35 (5.17%)	47 (6.51%)	199 (6.56%)
Genre conventions	42 (5.41%)	107 (12.46%)	65 (9.60%)	63 (8.73%)	277 (9.13%)
Information accuracy and clarity	93 (11.98%)	93 (10.83%)	80 (11.82%)	115 (15.93%)	381 (12.56%)
Lexis/wording	311 (40.08%)	281 (32.71%)	233 (34.42%)	305 (42.24%)	1130 (37.24%)
Morphosyntax	181(23.32%)	209 (24.33%)	176 (26.00%)	118 (16.34%)	684 (22.54%)
Overall quality	20 (2.58%)	24 (2.79%)	6 (0.89%)	7 (0.97%)	57 (1.88%)
Orthographic issues	56 (7.22%)	48 (5.59%)	52 (7.68%)	43 (5.96%)	199 (6.56%)
Total	776 (100%)	859 (100%)	677 (100%)	722 (100%)	3034 (100%)

*The numbers in the table indicate feedback frequency, with percentages in parentheses showing the proportion of each feedback content category provided by ChatGPT for each task and overall.

Table 5. Feedback content provided by teachers*.

Content	Task1	Task2	Task3	Task4	Overall
Coherence	11 (2.95%)	32 (7.98%)	25 (4.84%)	23 (4.37%)	91 (5.01%)
Cohesion	31 (8.31%)	18 (4.49%)	31 (6.00%)	29 (5.51%)	109 (6.00%)
Genre conventions	65 (17.43%)	63 (15.71%)	81 (15.67%)	71 (13.50%)	280 (15.42%)
Information accuracy and clarity	77 (20.64%)	102 (25.44%)	80 (15.47%)	115 (21.86%)	374 (20.58%)
Lexis/wording	49 (13.14%)	75 (18.70%)	100 (19.34%)	81 (15.40%)	305 (16.79%)
Morphosyntax	81 (21.72%)	34 (8.48%)	144 (27.85%)	133 (25.29%)	392 (21.57%)
Overall quality	13 (3.49%)	42 (10.47%)	10 (1.93%)	37 (7.03%)	102 (5.61%)
Orthographic issues	46 (12.33%)	35 (8.73%)	46 (8.90%)	37 (7.03%)	164 (9.03%)
Total	373 (100%)	401 (100%)	517 (100%)	526 (100%)	1817 (100%)

*The numbers in the table indicate feedback frequency, with percentages in parentheses showing the proportion of each feedback content category provided by human evaluators for each task and overall.

However, the quantity and distribution of feedback across different content categories differed between ChatGPT and teachers. ChatGPT provided significantly more feedback overall, while teachers offered a more balanced coverage of various content aspects (Figure 2). Notably, more than half of ChatGPT's feedback focused on lexical and morphosyntactic issues across the four tasks, with information accuracy and clarity (Info) being the third focus, constituting one-tenth of the feedback. In contrast, ChatGPT placed little emphasis on genre conventions (Genre) and text cohesion, with even less attention given to orthographic errors (Written) and text coherence. Teachers, on the other hand, provided a more even distribution of feedback on genre conventions, information accuracy, and clarity, lexical choices, and morphosyntactic issues, with

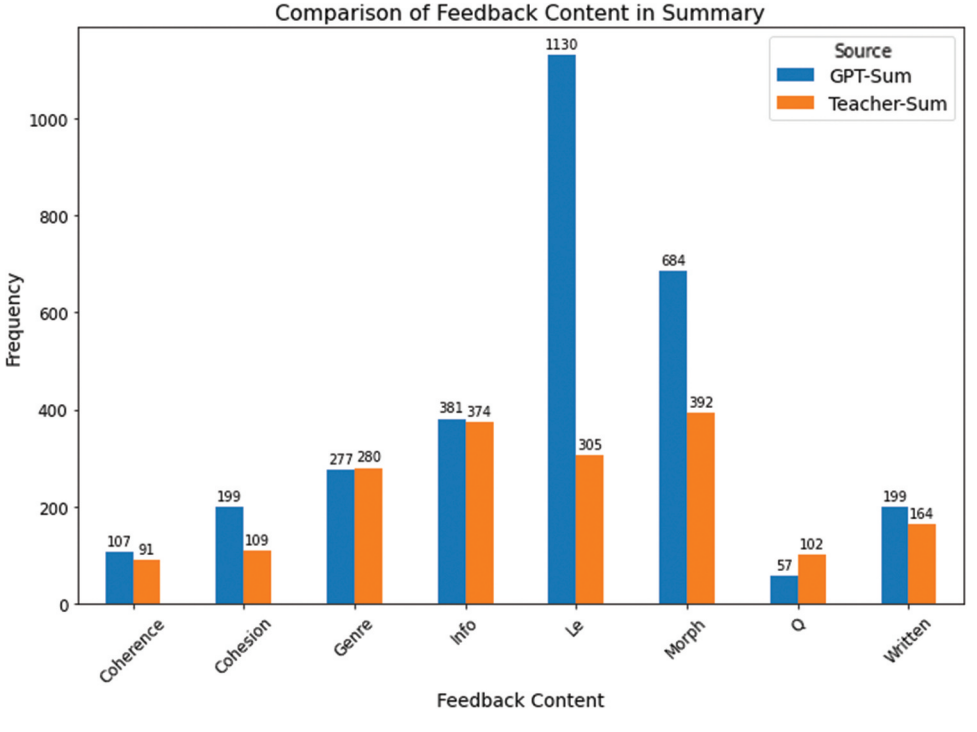

Figure 2. Comparison of feedback content provided by ChatGPT and teachers.

information and morphosyntax each accounting for over twenty percent of the feedback, and lexical and genre issues slightly exceeding fifteen percent.

4.1.2. Feedback strategy

A chi-square test of independence revealed a statistically significant association between the feedback strategy and the feedback provider (GPT or human) across all four tasks (Task 1: χ^2 (3, 1149) = 425.75, $p < 0.001$; Task 2: χ^2 (3, 1260) = 672.03, $p < 0.001$; Task 3: χ^2 (3, 1194) = 380.78, $p < 0.001$; Task 4: χ^2 (3, 1248) = 425.40, $p < 0.001$). A significant association was also found for the combined data across all four tasks (χ^2 (3, 4851) = 1881.23, $p < 0.001$).

Tables 6 and 7, together with Figure 3, demonstrate more differences than similarities in the feedback strategies employed by ChatGPT and human evaluators. ChatGPT's feedback was predominantly direct, accounting for over half of the total feedback, followed by general comments, which made up more than one-fourth of the feedback. However, indirect feedback was scarce in ChatGPT's responses. In contrast, human evaluators preferred indirect feedback, which identified problems in student translations without providing direct solutions. They might not offer general comments across every dimension for each student, resulting in a lower number of general comments compared to ChatGPT. Regarding explanatory feedback, while ChatGPT's feedback rarely included justifications, human evaluators were more inclined to provide explanations, regardless of whether their feedback was direct or indirect.

Table 6. Feedback strategy provided by ChatGPT*.

Strategy	Task1	Task2	Task3	Task4	Overall
Direct0	413 (53.22%)	441 (51.34%)	282 (41.65%)	419 (58.03%)	1555 (51.25%)
Direct1	3 (0.39%)	3 (0.35%)	6 (0.89%)	0 (0.00%)	12 (0.40%)
Indirect1	37 (4.77%)	6 (0.70%)	25 (3.69%)	30 (4.16%)	98 (3.23%)
Indirect2	2 (0.26%)	1 (0.12%)	3 (0.44%)	2 (0.28%)	8 (0.26%)
Indirect3	0 (0.00%)	5 (0.58%)	1 (0.15%)	1 (0.14%)	7 (0.23%)
Re0	124 (15.98%)	178 (20.72%)	147 (21.71%)	35 (4.85%)	484 (15.95%)
Re1	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Gen	197 (25.39%)	225 (26.19%)	213 (31.46%)	235 (32.55%)	870 (28.68%)
Total	776 (100%)	859 (100%)	677 (100%)	722 (100%)	3034 (100%)

*The numbers in the table indicate feedback frequency, with percentages in parentheses showing the proportion of each feedback strategy category provided by ChatGPT for each task and overall.

Table 7. Feedback strategy provided by teachers*.

Strategy	Task1	Task2	Task3	Task4	Overall
Direct0	35 (9.38%)	15 (3.74%)	96 (18.57%)	66 (12.55%)	212 (11.67%)
Direct1	22 (5.90%)	7 (1.75%)	35 (6.77%)	44 (8.37%)	108 (5.94%)
Indirect1	110 (29.49%)	105 (26.18%)	157 (30.37%)	153 (29.09%)	525 (28.89%)
Indirect2	39 (10.46%)	97 (24.19%)	77 (14.89%)	95 (18.06%)	308 (16.95%)
Indirect3	51 (13.67%)	21 (5.24%)	26 (5.03%)	43 (8.17%)	141 (7.76%)
Re0	3 (0.80%)	1 (0.25%)	3 (0.58%)	2 (0.38%)	9 (0.50%)
Re1	2 (0.54%)	4 (1.00%)	5 (0.97%)	3 (0.57%)	14 (0.77%)
Gen	111 (29.76%)	151 (37.66%)	118 (22.82%)	120 (22.81%)	500 (27.52%)
Total	373 (100.00%)	401 (100%)	517 (100%)	526 (100%)	1817 (100%)

*The numbers in the table indicate feedback frequency, with percentages in parentheses showing the proportion of each feedback strategy category provided by teachers for each task and overall.

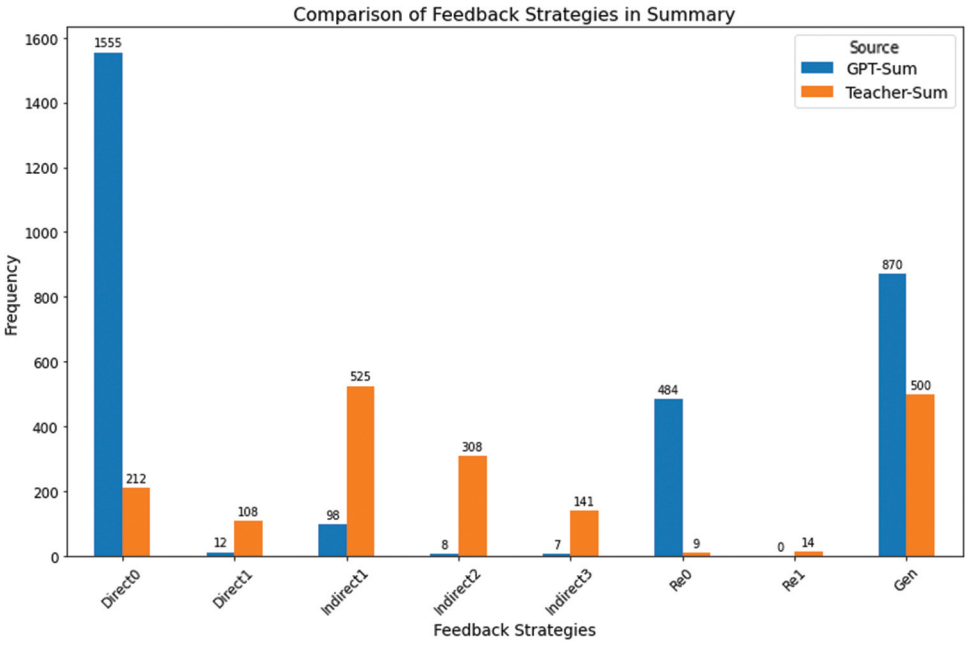

Figure 3. Comparison of feedback strategy provided by ChatGPT and teachers.

4.2. Students' revision behaviors

A chi-square test of independence was performed to investigate the relationship between students' revision behaviours and the feedback source across all four tasks. The results revealed a statistically significant association between these variables in each task (Task 1: χ^2 (5, 1353) = 188.25, $p < .001$; Task 2: χ^2 (5, 1337) = 108.92, $p < .001$; Task 3: χ^2 (5, 1253) = 87.75, $p < .001$; and Task 4: χ^2 (5, 1298) = 140.38, $p < .001$). The chi-square test also yielded a significant result when examining the combined data across all tasks, χ^2 (5, 5241) = 481.06, $p < .001$.

Table 8, Table 9, and Figure 4 show that, regardless of the feedback source, the most common student response was the absence of any revision aligned with the feedback provided. This non-revision behaviour accounted for nearly half of all observed behaviours, with students more likely to ignore feedback from ChatGPT across all four tasks. Apart from this, effective acceptance of feedback, where students incorporated suggestions without changes, was the second most prevalent behaviour. Such behaviours were more frequent with ChatGPT feedback than with teacher feedback, likely due to ChatGPT's substantial provision of direct translation references. In comparison, students were also more likely to make effective alterations and develop their own translations (i.e., effective substitutions) after considering teacher feedback rather than from ChatGPT. Regarding ineffective revisions, more were found when students responded to ChatGPT feedback. Furthermore, students often initiated revisions beyond the feedback's scope, with a higher occurrence of effective self-initiated revisions when interacting with teacher feedback. Concurrently, the incidence of ineffective self-initiated revisions was relatively low, irrespective of the feedback sources.

Table 8. Revision behaviours matched with ChatGPT feedback*.

Revision	Task 1	Task 2	Task 3	Task 4	Total
No change	440 (50.75%)	489 (54.39%)	483 (69.30%)	500 (67.29%)	1912(59.64%)
Effective acceptance	279 (32.18%)	266 (29.59%)	139 (19.94%)	134 (18.03%)	818 (25.51%)
Effective substitution	15 (1.73%)	16 (1.78%)	30 (4.30%)	12 (1.62%)	73 (2.28%)
Ineffective revision related to feedback	42 (4.84%)	88 (9.79%)	25 (3.59%)	76 (10.23%)	231 (7.21%)
Effective self-initiated revision	49 (5.65%)	24 (2.67%)	13 (1.87%)	16 (2.15%)	102 (3.18%)
Ineffective self-initiated revision	42 (4.84%)	16 (1.78%)	7 (1.00%)	5 (0.67%)	70 (2.18%)
Total	867 (100%)	899 (100%)	697 (100%)	743 (100%)	3206 (100%)

*The numbers in the table indicate the frequency of revision behaviours, with percentages in parentheses showing the proportion of each type relative to the total revisions per task and overall.

Table 9. Revision behaviours matched with teacher feedback*.

Revision	Task 1	Task 2	Task 3	Task 4	Total
No Change	154 (31.69%)	238 (54.34%)	289 (51.98%)	321 (57.84%)	1002 (49.24%)
Effective Acceptance	119 (24.49%)	84 (19.18%)	124 (22.30%)	90 (16.22%)	417 (20.49%)
Effective Substitution	87 (17.90%)	55 (12.56%)	97 (17.45%)	102 (18.38%)	341 (16.76%)
Ineffective revision related to feedback	13 (2.67%)	24 (5.48%)	7 (1.26%)	13 (2.34%)	57 (2.80%)
Effective Self-initiated Revision	85 (17.49%)	36 (8.22%)	31 (5.58%)	23 (4.14%)	175 (8.60%)
Ineffective Self-initiated Revision	28 (5.76%)	1 (0.23%)	8 (1.44%)	6 (1.08%)	43 (2.11%)
Total	486 (100%)	438 (100%)	556 (100%)	555 (100%)	2035 (100%)

*The numbers in the table indicate the frequency of revision behaviours, with percentages in parentheses representing the proportion of each type relative to the total revision behaviours per task and overall.

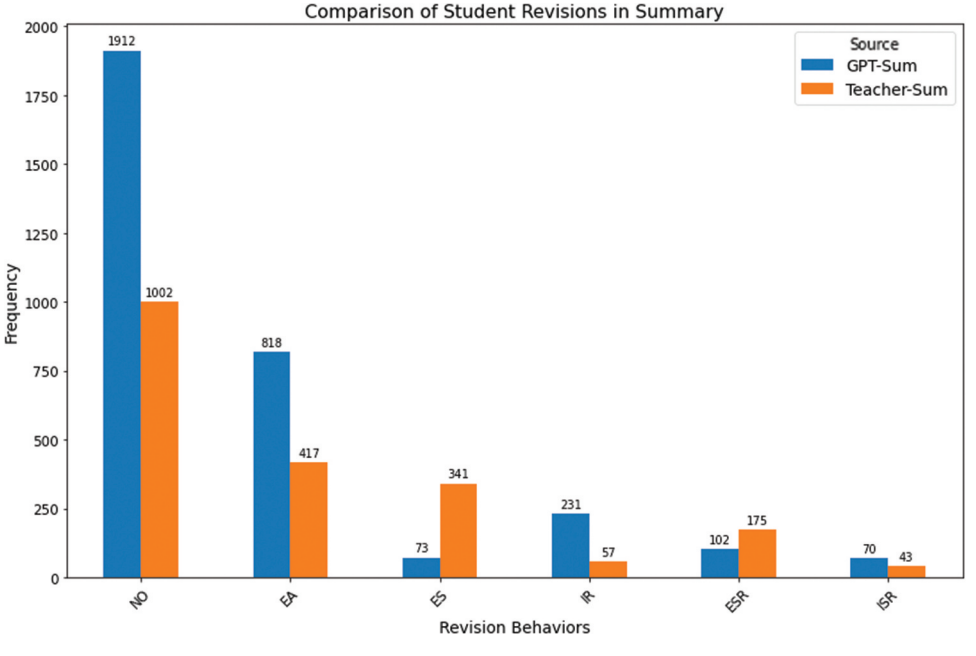

Figure 4. Comparison of student revision behaviors.

4.3. Students' perceptions of ChatGPT and teacher feedback

The thematic analysis elucidates students' perceptions of ChatGPT and teacher feedback by categorising their responses into four main themes: (1) quality and reliability, which assesses whether the feedback is accurate, helpful, and trustworthy; (2) comprehensibility, evaluating how easily the feedback can be understood; and (3) affective connection, describing their emotional reactions to the feedback.

4.3.1. Quality and reliability

Regarding the quality of feedback, all participants expressed satisfaction with teacher feedback. According to Q, teacher feedback was 'quite to the point' in identifying issues within their translations. None of the participants reported dissatisfaction with the quality of teacher feedback, and all expressed an almost 'unconditional trust' in it. Participant F highlighted that teachers were consistently experienced and knowledgeable, while Q emphasised the reliability of teacher feedback by referring to it as an 'authority'. Both Q and N noted that they would accept teachers' suggestions even when their own opinions differed.

In contrast, all interviewees found the overall quality of ChatGPT feedback acceptable but not up to their expectations. They were particularly satisfied with ChatGPT's variability in wording, sentence structure, and attention to grammatical details. For example, students appreciated ChatGPT's ability to effectively handle the 'prose-style and flexible expressions' of the travel narrative in Task 3. Participant Q emphasised that, despite difficulties in capturing the essence of Task 3, ChatGPT's reference translation vividly depicted the landscape without being overly influenced by the original Chinese text. Students also valued ChatGPT's proficiency in dealing with long sentences, with N appreciating that ChatGPT's modification of sentence structure could sometimes contribute to better coherence. Another praised aspect was ChatGPT's attention to grammatical details, such as tense, articles, and prepositions, which were frequently overlooked by teachers in other translation tasks (mentioned by Q and N).

In contrast, students expressed greater reservations about the quality and reliability of ChatGPT's feedback. While they initially trusted ChatGPT and attributed high credibility to its advanced technology, their confidence diminished after encountering errors in its feedback. For instance, participants L and N criticised ChatGPT for making unnecessary revisions. L questioned ChatGPT's ability to provide tailored feedback, asking, 'Has ChatGPT really examined my work, or does it simply present similar feedback to everyone?' Furthermore, although students acknowledged that ChatGPT's feedback could adapt to different genre conventions, they found its performance relatively unsatisfactory when addressing political texts, which often include culturally or politically specific expressions. L complained that ChatGPT's reference translation adhered too closely to the syntax and lexical choices of the Chinese source text, suggesting that there should have been modifications to better cater to the target readers. Participant N emphasised the importance of accurately conveying China's perspective in culturally and politically charged texts, a task that ChatGPT may struggle to manage effectively. Additionally, students highlighted issues with ChatGPT's feedback on terminology translation. Participant Q remarked that 'there are commonly accepted ways of translating certain expressions' that ChatGPT failed to adhere to. However, N expressed continued trust in

ChatGPT's feedback on grammatical issues, attributing this reliability to its training on a large corpus of texts, which likely enhanced its accuracy in grammar usage.

4.3.2. Comprehensibility of feedback

Students generally found teacher feedback more comprehensible than ChatGPT feedback. One reason for this difference is the distinct language features of teacher and ChatGPT feedback. N mentioned that teachers used 'simple words', while the wording of ChatGPT was more 'linguistic and professional', making it relatively more difficult to understand.

Another reason lies in the feedback strategies. Teachers provided more indirect feedback by identifying translation issues and occasionally offering explanations, but they seldom provided direct references. Q regarded this kind of feedback as beneficial for understanding their translation problems and encouraging independent thinking and learning. In comparison, students criticised ChatGPT feedback as being either 'too general' or 'too direct', typically consisting of detailed direct translation references followed by general comments. Students had to make comparisons between their own translations and the ChatGPT versions to identify their problems. F, for instance, expressed reluctance to spend much time on this comparison. Moreover, ChatGPT rarely provided the rationale behind the feedback. Regarding direct reference translation, X commented, 'Sometimes my understanding of the source text differed from that of ChatGPT. When I was uncertain, I struggled to determine which interpretation was correct'. Concerns were also raised about the potential over-reliance on ChatGPT without understanding the reasons for revision (expressed by X and N in the interview). In terms of general feedback from ChatGPT, they found them too vague to be actionable.

While some students found it challenging to comprehend ChatGPT feedback, they also saw it as a learning opportunity. Q, N, and W observed that comparing the varied expressions and sentence structures in ChatGPT's reference translations encouraged them to view their own translations from different perspectives, which W described as 'the sparkles of different ideas'. W noted that ChatGPT's feedback often appeared neutral and objective, differing from human feedback. For instance, Q recounted a case where ChatGPT changed her translation of '正确的道路' from 'the right way' to 'the correct way'. Initially, she was puzzled by this change. However, upon further reflection, she considered that the modification might have been made to avoid confusing the original meaning with the term 'right wing'. Similarly, N attempted to analyse the translation strategies in ChatGPT's reference translation. Furthermore, W mentioned that the general feedback from ChatGPT encouraged her to critically evaluate her translations, promoting self-reflection rather than passively accepting external feedback.

4.3.3. Affective connection

In addition, the affective connection students experienced varies between teacher feedback and ChatGPT feedback. N appreciated teachers' conscientiousness in providing feedback, which fostered a strong emotional bond and acted as a source of encouragement. In contrast, although ChatGPT often provided positive feedback, it failed to resonate emotionally with students. N described this feedback as 'overtly positive' and 'unhelpful', expressing that even positive remarks from ChatGPT did not 'genuinely' convince her of the quality of her work.

This discrepancy in the affective connection also influenced how students responded to it. Teacher feedback, though sometimes harsh, was more likely to motivate students to refine their work, whereas the generally positive nature of ChatGPT's feedback made students feel uneasy and sometimes under-motivated. Moreover, authorship concerns and AI replacement anxiety influenced students' willingness to accept ChatGPT's reference translations. For instance, Q expressed a desire to 'maintain the originality of her translation rather than simply copying AI suggestions', while multiple participants (F, Q, X, and W) voiced concerns about 'being replaced by AI.' As a result, they often felt compelled to develop their own revisions, although they may have opted not to make changes if they could not devise new strategies.

5. Discussion

5.1. Comparison between ChatGPT and teacher feedback

Both similarities and differences are found regarding the content and strategies of ChatGPT and teacher feedback. As for similarities, both ChatGPT and teachers prioritised similar aspects of feedback content, which indicates ChatGPT's capability to identify critical issues in student translations. The finding also corroborates L2 writing research, which found ChatGPT to achieve good consistency with human evaluators (Lu et al. 2024; Mizumoto and Eguchi 2023).

However, more differences were found than similarities. Overall, ChatGPT generated a greater amount of feedback than teachers, which echoes the findings of Guo and Wang (2024) in the EFL writing context. While this highlights ChatGPT's efficiency to produce feedback in seconds compared to the hours or even days teachers may require, it also leads to potential cognitive overload for students. The extensive feedback can be overwhelming, making it difficult for students to process and effectively use the feedback (c.f. Bitchener and Storch 2016; Lee 2019). For instance, participant F noted that she bypassed the critical analysis of feedback, a key step in learning from it. Moreover, learning is less likely to occur when feedback targets only the trivial issues in students' work (Lee 2019). This is particularly problematic because the ultimate aim of feedback is to help learners become competent translators, not to overwhelm them with detailed revisions.

The second notable difference between ChatGPT and teacher feedback is the focus of content. Contrary to Guo and Wang (2024), who found that ChatGPT 3.5 provided balanced feedback across various writing dimensions, this study revealed that teachers offered a more thorough assessment across different translation dimensions. ChatGPT 3.5, however, tended to focus on lexical and morphosyntactic issues, influenced by its task-specific nature in this translation study. In this study, ChatGPT offered reference translations containing changes primarily in lexical and morphosyntactic areas. Such nature of ChatGPT's feedback also influenced students' perceptions of its usefulness.

Feedback strategies also differ markedly between human evaluators and ChatGPT. Consistent with prior research (Guo and Wang 2024; Lu et al. 2024), ChatGPT's feedback was often overly general or directly corrective, similar to novice evaluators observed in L2 contexts (Li et al. 2023; Yu et al. 2020). This tendency may stem from ChatGPT's 'inherent limitations in logical reasoning' (Su et al. 2023, 10; Lu et al. 2024, 625). In contrast, teachers preferred to use indirect feedback that explicitly identified problems with some explanations

but did not provide immediate solutions. The distinct strategies employed by teachers and ChatGPT when providing feedback contribute to the contrasting attitudes students hold regarding the quality and usefulness of the feedback from these two sources, as observed in this study. Research indicates that the most effective feedback possesses two key features: specificity and clear guidance for improving knowledge and skills (Dawson et al. 2019, 32), which are characteristics of the feedback typically provided by teachers. Conversely, the highly directed feedback or specific improvement suggestions offered by ChatGPT may lead to changes that are limited to particular tasks. Such feedback has not been recognised as crucial for effective learning by either teachers or students (Dawson et al. 2019, 32).

Furthermore, adherence to the multifaceted criteria set out in the assessment rubrics led ChatGPT to generate a larger number of general comments than human evaluators. This observation aligns with that of Steiss et al. (2024), who reported that ChatGPT demonstrated better compliance with criteria than teachers. It should be noted that although general comments might not offer immediate practical revisions, they provide students with a global understanding of their strengths and weaknesses, potentially encouraging more proactive revisions in the future (Ferris 1997, 333).

5.2. Student perception and adoption of feedback

This study, echoing L2 writing research (Lu et al. 2024), reveals that students are more likely to adopt feedback from teachers than from ChatGPT because of its perceived reliability and usefulness. For instance, students demonstrated a strong trust in teacher feedback, often accepting it even when they held differing opinions. In contrast, while students were willing to consider ChatGPT's recommendations concerning lexical choices, syntactic structures, and grammatical issues, they tended to adhere to their own translations in other aspects due to concerns about its reliability.

Another possible factor influencing student behaviours is how well they understand the feedback. According to the cognitive processing model of feedback, for revision and learning to happen, students must first notice the knowledge gaps in the feedback and then comprehend the feedback provided (Bitchener and Storch 2016, 19). Some students viewed the challenges in understanding ChatGPT's feedback as an opportunity for learning and reflection, which echoes the finding of Gozali et al. (2024). This motivation and positive perceptions of feedback might stem from their growth mindset – the belief that they can develop their skills through effort and by learning from difficulties (Yao et al. 2024, 2).

Moreover, the emotional responses to the feedback provider play a crucial role in how feedback is received and acted upon. Feedback is generally more effective when there is a supportive emotional connection between the provider and the receiver (Steen-Utheim and Wittek 2017, 28). In contrast to previous research where students preferred positive feedback (Neunzig and Tanqueiro 2005), the current research found that students reacted positively to teachers' critical feedback and expressed dissatisfaction with ChatGPT's overly positive feedback. This reaction might be due to the trust and rapport established between teachers and students, enabling students to manage emotional responses constructively and engage seriously with critical feedback (Carless and Boud 2018, 1318). Furthermore, concerns about cheating and fears of being replaced by technology might make students hesitant to fully embrace ChatGPT's reference

translations. This complex interplay of trust, emotional connection, and feedback comprehension underscores the multifaceted nature of feedback adoption in translation education.

5.3. Pedagogical implications

Despite its limitations, ChatGPT can effectively complement teacher feedback in translation education if applied judiciously. Firstly, considering that timeliness is essential for effective feedback (Dawson et al. 2019, 30), the high efficiency of ChatGPT can greatly reduce the time between the initial draft and subsequent revisions, providing a timely and quality alternative when immediate feedback from teachers is unavailable (Steiss et al. 2024, 13). Moreover, students expressed a desire for further interactions with teachers following their revisions even though this may not always be feasible. While only a single ChatGPT response was used for experimental purposes in this study, its interactive capabilities (Su et al. 2023, 2) could, in future translation education settings, support a more dialogic learning experience, as envisioned by Washbourne (2014).

Second, as ChatGPT and teachers often focus on different aspects of the translation and employ various strategies in providing feedback, they can complement each other, offering a more thorough review of students' translations. Moreover, when feedback from both sources overlaps, it can reinforce students' awareness and understanding of the key issues involved, helping them recognise the importance of that particular issue and better understand how to address it in their translations (Lu et al. 2024, 625).

Third, although teacher feedback is generally reliable, an over-reliance on it might limit opportunities for critical thinking. The vague and general nature of ChatGPT feedback, along with students' scepticism towards its quality and reliability, can prompt students to think independently and critically in the learning process. This evaluative judgement is a crucial component of feedback literacy (Carless and Boud 2018; Lu et al. 2024). During the process, students can engage in deeper cognitive processing of feedback by comparing, synthesising, and evaluating the output from ChatGPT (c.f. Bitchener and Storch 2016; Carless and Boud 2018). This process might also consolidate students' understanding of the criteria of good translation as well as the strengths and drawbacks of ChatGPT feedback, enabling them to make more informed judgements in the future (Lu et al. 2024, 626). By repeatedly engaging in the critical evaluation of ChatGPT's feedback, students can effectively transfer the knowledge and skills they acquire during this process into their long-term memory (Bitchener and Storch 2016, 25).

Moreover, effective use of ChatGPT feedback in translation education requires teacher guidance. While integrating ChatGPT feedback with teacher feedback can improve feedback efficiency, the prerequisite is to ensure that students are well-informed about the source of the feedback. As defined by Ng et al. (2021, 5), AI literacy encompasses essential competencies such as understanding and recognising AI, utilising and applying it effectively, and evaluating and creating AI technologies. Therefore, it is crucial to first introduce students to the underlying technologies of ChatGPT, as well as its strengths and weaknesses (Ng et al. 2021; Su et al. 2023). Furthermore, responsible use of ChatGPT feedback necessitates awareness of potential ethical issues, such as plagiarism risks if ChatGPT's suggestions are accepted uncritically (Su et al. 2023, 9). Teachers can aid students in generating useful feedback from ChatGPT by providing effective sample prompts, teaching prompt engineering techniques

(Ekin 2023; Guo and Wang 2024, 20), and guiding students on how to critically and effectively engage with ChatGPT feedback. In the contexts where ChatGPT provides reference translations, teachers could use sample translations to illustrate common issues associated with AI-generated reference translations and prepare students for critical evaluation and adoption of translation feedback generated by ChatGPT.

6. Conclusion

This study expands the existing body of literature by examining the characteristics of feedback provided by ChatGPT and human evaluators in the context of translation teaching, as well as how translation students perceive and utilise the two types of feedback. Based on these findings, the study proposes potential approaches to using ChatGPT feedback as a supplement to teacher feedback in translation education.

Despite the contributions, several limitations should be acknowledged. First, the present study utilised the GPT-3.5 model, which was the most accessible and prevalent model at the time of the experiment. It is important to note that due to the rapid advancements in generative AI technologies, newer models such as GPT-4o and Claude-3 have been developed, which may yield different results. Future research could explore the capabilities of these more advanced AI models in providing translation feedback. Second, this study employed a fixed prompt to generate ChatGPT's feedback on student translations. Although this prompt was rigorously developed and refined by researchers, changes in the wording could result in different outcomes (OpenAI 2022). In future educational settings, teachers could enhance the effectiveness of ChatGPT feedback by providing detailed information on task specifics, relevant terminology, student proficiency levels, and desired feedback characteristics, allowing for more precisely tailored feedback to meet student needs. Moreover, since participants noted that ChatGPT's performance varied across different text types, future research could investigate this by conducting more controlled studies to explore how ChatGPT feedback changes with different text types, thereby further enriching our understanding of its potential application in translation education.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Hong Kong Polytechnic University [TDG22-25/VTL-6].

ORCID

Yanfang Su http://orcid.org/0000-0002-4526-4187
Simin Xu http://orcid.org/0009-0000-9463-5557
Kanglong Liu http://orcid.org/0000-0003-3962-2563

References

Behr, M. 2015. "How to Back the Students - Quality, Assessment & Feedback." To Know How to Suggest: Approaches to Teaching Conference Interpreting, edited by D. Andres and M. Behr. Berlin: 201–217. Frank & Timme.

Bitchener, J., and N. Storch. 2016. *Written Corrective Feedback for L2 Development*. Bristol: Multilingual Matters.

Bruton, A. 2007. "Vocabulary Learning from Dictionary Referencing and Language Feedback in EFL Translational Writing." *Language Teaching Research* 11 (4): 413–431. https://doi.org/10.1177/1362168807080961.

Cao, S., and L. Zhong. 2023. "Exploring the Effectiveness of ChatGPT-Based Feedback Compared with Teacher Feedback and Self-Feedback: Evidence from Chinese to English Translation." *Computing Research Repository* arXiv:2309.01645. https://arxiv.org/abs/2309.01645.

Carless, D., and D. Boud. 2018. "The Development of Student Feedback Literacy: Enabling Uptake of Feedback." *Assessment & Evaluation in Higher Education* 43 (8): 1315–1325. https://doi.org/10.1080/02602938.2018.1463354.

Cohen, L., L. Manion, and K. Morrison. 2011. *Research Methods in Education*. London: Routledge.

Dawson, P., M. Henderson, P. Mahoney, M. Phillips, T. Ryan, D. Boud, and E. Molloy. 2019. "What Makes for Effective Feedback: Staff and Student Perspectives." *Assessment & Evaluation in Higher Education* 44 (1): 25–36. https://doi.org/10.1080/02602938.2018.1467877.

Ekin, S. 2023. "Prompt Engineering for ChatGPT: A Quick Guide to Techniques, Tips, and Best Practices." *TechRxiv*. https://doi.org/10.36227/techrxiv.22683919.v2.

Ellis, R. 2009. "A Typology of Written Corrective Feedback Types." *ELT Journal* 63 (2): 97–107. https://doi.org/10.1093/elt/ccn023.

Ferris, D. R. 1997. "The Influence of Teacher Commentary on Student Revision." *TESOL Quarterly* 31 (2): 315–339. https://doi.org/10.2307/3588049.

Floridi, L. 2023. "AI as Agency without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models." *Philosophy & Technology* 36 (1): 15. https://doi.org/10.1007/s13347-023-00621-y.

Giray, L. 2023. "Prompt Engineering with ChatGPT: A Guide for Academic Writers." *Annals of Biomedical Engineering* 51 (12): 2629–2633. https://doi.org/10.1007/s10439-023-03272-4.

Gozali, I., A. R. T. Wijaya, A. Lie, B. Y. Cahyono, and N. Suryati. 2024. "Leveraging the Potential of ChatGPT as an Automated Writing Evaluation (AWE) Tool: Students' Feedback Literacy Development and AWE Tools Integration Framework." *The JALT CALL Journal* 20 (1): 1–22. https://doi.org/10.29140/jaltcall.v20n1.1200.

Guo, K., and D. Wang. 2024. "To Resist it or to Embrace It? Examining ChatGPT's Potential to Support Teacher Feedback in EFL Writing." *Education and Information Technologies* 29 (7): 8435–8463. https://doi.org/10.1007/s10639-023-12146-0.

Han, C., and X. Lu. 2023. "Can Automated Machine Translation Evaluation Metrics Be Used to Assess Students' Interpretation in the Language Learning Classroom?" *Computer Assisted Language Learning* 36 (5–6): 1064–1087. https://doi.org/10.1080/09588221.2021.1968915.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. https://doi.org/10.3102/003465430298487.

Hurtado Albir, A. 2015. "The Acquisition of Translation Competence. Competences, Tasks, and Assessment in Translator Training." *Meta* 60 (2): 256–280. https://doi.org/10.7202/1032857ar.

Hurtado Albir, A., and S. Pavani. 2018. "An Empirical Study on Multidimensional Summative Assessment in Translation Teaching." *Interpreter and Translator Trainer* 12 (1): 25–47. https://doi.org/10.1080/1750399X.2017.1420131.

Korol, T. 2022. "Digital Teacher Feedback as a Translation Assessment Tool." *Journal of Teaching English for Specific and Academic Purposes*: 575–586. https://doi.org/10.22190/JTESAP2104575K.

Lee, I. 2019. "Teacher Written Corrective Feedback: Less is More." *Language Teaching* 52 (4): 524–536. https://doi.org/10.1017/S0261444819000247.

Li, M., S. Yu, P. Mak, and C. Liu. 2023. "Exploring the Efficacy of Peer Assessment in University Translation Classrooms." *Interpreter and Translator Trainer* 17 (4): 585–609. https://doi.org/10.1080/1750399X.2023.2236920.

Link, S., M. Mehrzad, and M. Rahimi. 2022. "Impact of Automated Writing Evaluation on Teacher Feedback, Student Revision, and Writing Improvement." *Computer Assisted Language Learning* 35 (4): 605–634. https://doi.org/10.1080/09588221.2020.1743323.

Lu, Q., Y. Yao, L. Xiao, M. Yuan, J. Wang, and X. Zhu. 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.

Mizumoto, A., and M. Eguchi. 2023. "Exploring the Potential of Using an AI Language Model for Automated Essay Scoring." *Research Methods in Applied Linguistics* 2 (2): 100050. https://doi.org/10.1016/j.rmal.2023.100050.

Neunzig, W., and H. Tanqueiro. 2005. "Teacher Feedback in Online Education for Trainee Translators." *Meta* 50 (4): 1–10. https://doi.org/10.7202/019873ar.

Ng, D. T. K., J. K. L. Leung, S. K. W. Chu, and M. S. Qiao. 2021. "Conceptualizing AI Literacy: An Exploratory Review." *Computers and Education: Artificial Intelligence* 2:100041. https://doi.org/10.1016/j.caeai.2021.100041.

Nowell, L. S., J. M. Norris, D. E. White, and N. J. Moules. 2017. "Thematic Analysis: Striving to Meet the Trustworthiness Criteria." *International Journal of Qualitative Methods* 16 (1): 1–13. https://doi.org/10.1177/1609406917733847.

OpenAI. 2022. "Introducing ChatGPT." OpenAi. November 30, 2022. https://openai.com/blog/chatgpt.

Patton, M. Q. 2014. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. 4th ed. California: Sage Publications.

Shadiev, R., and Y. Feng. 2023. "Using Automated Corrective Feedback Tools in Language Learning: A Review Study." *Interactive Learning Environments* 32 (6): 2538–2566. https://doi.org/10.1080/10494820.2022.2153145.

Sofyan, R., and D. Y. Nasution. 2023. "Using Grammarly as a Tool to Improve the Students' Grammatical Accuracy in Translation." *English Language and Literature* 12 (2): 243–254. http://ejournal.unp.ac.id/index.php/jell.

Steen-Utheim, A., and A. L. Wittek. 2017. "Dialogic Feedback and Potentialities for Student Learning." *Learning, Culture & Social Interaction* 15:18–30. https://doi.org/10.1016/j.lcsi.2017.06.002.

Steiss, J., T. Tate, S. Graham, J. Cruz, M. Hebert, J. Wang, Y. Moon, W. Tseng, M. Warschauer, and C. B. Olson. 2024. "Comparing the Quality of Human and ChatGPT Feedback of Students' Writing." *Learning & Instruction* 91:101894. https://doi.org/10.1016/j.learninstruc.2024.101894.

Su, Y., Y. Lin, and C. Lai. 2023. "Collaborating with ChatGPT in Argumentative Writing Classrooms." *Assessing Writing* 57:100752. https://doi.org/10.1016/j.asw.2023.100752.

Sun, S. 2015. "Measuring Translation Difficulty: Theoretical and Methodological Considerations." *Across Languages and Cultures* 16 (1): 29–54. https://doi.org/10.1556/084.2015.16.1.2.

Tabrizi, A. R. N., and A. P. Etemad. 2021. "Check the Grammar of Your Translation: The Usefulness of Whitesmoke as an Automated Feedback Program." *Digital Scholarship in the Humanities* 36 (2): 411–420. https://doi.org/10.1093/llc/fqaa013.

Washbourne, K. 2014. "Beyond Error Marking: Written Corrective Feedback for a Dialogic Pedagogy in Translator Training." *Interpreter and Translator Trainer* 8 (2): 240–256. https://doi.org/10.1080/1750399X.2014.908554.

Yao, Y., X. Zhu, and J. Zhan. 2024. "The Associations Among Growth Mindsets, the Ideal L2 Writing Self, and L2 Writing Enjoyment and Their Impacts on L2 English Writing Performance: A Gender Difference Perspective." *Assessing Writing* 60:100832. https://doi.org/10.1016/j.asw.2024.100832.

Yu, S., Y. Zhang, Y. Zheng, and Z. Lin. 2020. "Written Corrective Feedback Strategies in English-Chinese Translation Classrooms." *The Asia-Pacific Education Researcher* 29 (2): 101–111. https://doi.org/10.1007/s40299-019-00456-2.

Appendices

Appendix A A sample ChatGPT feedback

- Reference Translation:
- Due to the need for both China and the United States to improve their relationship, high-level meetings between the two countries have been frequent in a short period of time.
- The marathon 12-hour meeting between Wang Yi and Sullivan covered topics such as Taiwan, Asia-Pacific, and international affairs.
- The discussions were practical and in-depth, and both sides agreed to have more high-level exchanges and consultations in the coming months regarding key areas of concern.
- When Blinken met with Han Zheng, he was frank and said, We have an opportunity to move forward based on recent high-level contacts between the two countries, which is a good thing.
- He emphasized that face-to-face diplomacy is the best way to handle differences, and he hoped that both sides would responsibly manage differences, promote cooperation, and stabilize the development of U.S.-China relations.
- Comments:
- The translation accurately conveys the meaning and tone of the original text, and it is well-organized and coherent.
- The vocabulary and morphosyntax are appropriate, and the conventions of written language are followed. The only minor issue is the use of the word marathon to describe the meeting between Wang Yi and Sullivan, which may not be the best choice of words.
- The level of communication of the target text is good, and it effectively conveys the message to the target audience.
- Overall, it is a high-quality translation.

Appendix B A sample teacher feedback

- Detailed feedback:
- The relations between China and the US can also be translated as ”Sino-US relations”.
- “逾12小时” is different from 12 hours.
- The target readers may not know who is “Wang Yi” and “Han Zheng”.
- “other affairs concerning the world” can be translated into high-level dialogues.
- “hoping both of us”: unclear “us”.
- “booster” is a noun, should be “boost”.
- Overall Feedback:
- EXPRESSION OF THE MEANING OF THE ORIGINAL TEXT: The translation effectively conveys the meaning of the original text with generally accurate information.
- COMPOSITION IN THE TARGET LANGUAGE: The translation has some instances where the news style could be improved to better align with the target language context and conventions. The appropriateness and richness of vocabulary can be improved. The cohesion could be improved with better use of connectors and referential elements.
- LEVEL OF COMMUNICATION OF THE TARGET TEXT: The target text demonstrates good quality and adheres to most of the genre’s conventions. It is generally appropriate for the translation’s purpose and target audience, although some improvements can be made to ensure better alignment.