Original Research

SAGE Open
July-September 2025: 1–18

DOI: 10.1177/21582440251369204
journals.sagepub.com/home/sgo
Sage

Exploring the Efficacy of ChatGPT-Based Feedback Compared With Teacher Feedback and Self-Feedback: Evidence From Chinese-English Translation

Siyi Cao[1,2] and Tongquan Zhou[1]

Abstract

ChatGPT, a cutting-edge AI-powered Chatbot, can quickly generate responses to given commands. While ChatGPT was reported to have the capacity to deliver useful feedback, it is still unclear about its effectiveness compared with conventional feedback approaches, such as self-feedback (SF) and teacher feedback (TF). To address this issue, this study compared the revised Chinese to English translation texts produced by 45 Chinese Master of Translation and Interpretation (MTI) students, who learned English as a Second Language (ESL), based on three feedback types (i.e., SF, TF, and ChatGPT feedback). The data was analyzed using BLEU score to gauge the overall translation quality as well as Coh-Metrix to examine linguistic features across three dimensions: lexicon, syntax, and cohesion. The findings revealed that SF and TF-guided translation texts surpassed those with ChatGPT feedback, as indicated by the BLEU score. In terms of linguistic features, ChatGPT feedback demonstrated superiority, particularly in enhancing lexical capability and referential cohesion in the translation texts. However, SF and TF proved more effective in developing syntax-related skills, as they addressed instances of incorrect usage of the passive voice. These diverse outcomes indicate ChatGPT's potential as a supplementary resource, complementing traditional teacher-led methods in translation practice.

Plain language summary

Assessing the Efficacy of ChatGPT-based Feedback in Chinese to English Translation: A Comparative Study with Teacher and Self-Feedback

Feedback plays a crucial role in the process of learning English as a second language (ESL), as it supports student motivation and achievement. ChatGPT, a cutting-edge AI-powered Chatbot, can aid ESL learners by providing instant and personalized feedback proved by theoretical studies. However, it is still unclear about its effectiveness compared with conventional feedback approaches, such as teacher feedback (TF) and self-feedback (SF). The aim of the present study is to compare the quality of Chinese to English translation texts produced by Chinese Master of Translation and Interpretation (MTI) students based on three feedback types (i.e., ChatGPT-based feedback, TF, and SF). A total of 135 translation texts were collected from 45 MTI participants, each subjected to three rounds

[1]Southeast University, Nanjing, China
[2]Hong Kong Polytechnic University, Kowloon, Hong Kong

Corresponding Authors:
Siyi Cao, School of Foreign Languages, Southeast University, Jiulonghu Campus, Nanjing 211189, China.
Email: siyi.c@nuaa.edu.cn

Tongquan Zhou, School of Foreign Languages, Southeast University, Jiulonghu Campus, Nanjing 211189, China.
Email: zhoutongquan@126.com

Data Availability Statement included at the end of the article

of feedback-driven revisions. Our analysis framework encompassed two main aspects: the overall translation quality and the linguistic dimensions. The findings contribute to the ongoing discussion about the role of AI by highlighting the specific strengths and weaknesses of ChatGPT in translator training. From a theoretical standpoint, the findings illuminate the limitations of AI in handling the complexities of human linguistic abilities, particularly in the realm of syntax, suggesting a need for further development in this area. On a practical level, our study indicates ChatGPT's potential as a supplementary resource, complementing traditional teacher-led methods in translation practice.

Keywords
self-feedback, teacher feedback, ChatGPT feedback, translation texts, MTI students

Introduction

Feedback plays a crucial role in the process of learning English as a Second Language (ESL; Cao et al., 2022; Hyland & Hyland, 2006), as it fuels student motivation and achievement (Cauley & McMillan, 2010). Different types of feedback exist, such as self-feedback (SF), teacher feedback (TF), and computer-generated feedback (CF; Hattie & Timperley, 2007; Lipnevich & Smith, 2022). However, conflicting results have emerged from previous research when comparing the effectiveness of these feedback types. Some studies indicated that TF was superior to CF in identifying grammatical errors and improving overall writing quality (Kaivanpanah et al., 2020; Park, 2019). Conversely, other scholars argued that CF surpassed TF in reducing grammatical errors and positively impacting ESL learners' writing ability (Hernández Puertas, 2018; Sistani & Tabatabaei, 2023). Moreover, TF and CF could eventually transition into SF (Lipnevich & Smith, 2022). Given the diversity of opinions and findings, further research is necessary to determine the optimal feedback approach for ESL learners.

ChatGPT, developed by OpenAI, is an AI-powered chatbot that has been hailed as a game-changer for ESL learners. While qualitative studies show its potential for ESL learning (Kasneci et al., 2023; Kuhail et al., 2023), experimental research on the effectiveness of its generated feedback is still scarce. To bridge this research gap, the present study aims to assess the impact of ChatGPT feedback in comparison to TF and SF in terms of the translation performance of advanced ESL learners, specifically Master of Translation and Interpreting (MTI) students in China. This comparative analysis would examine the overall translation quality (based on BLEU scores) as well as linguistic features such as lexicon, syntax, and cohesion in the students' revised translation texts, focusing on the three feedback types. The findings will shed light on the advantages and disadvantages of using AI Chatbot for feedback in the context of translation practice.

Self-Feedback Versus Teacher Feedback Versus Computer-Generated Feedback

Self-feedback (SF), as a self-regulated learning practice, often involves learners detecting and correcting their own mistakes based on prior knowledge and experience. It is highly recommended for practical use in ESL classrooms, as it provides opportunities for students to critically evaluate their texts and cultivate meta-awareness and autonomy in learning (Cahyono & Rosyida, 2016). Additionally, SF can increase student motivation and active participation in second-language writing, as well as create a self-paced learning environment (Miranty & Widiati, 2021; Yu, Jiang, & Zhou, 2020). However, SF may prove counterproductive if students' language proficiency is insufficient for independently identifying and rectifying all errors (Srichanyachon, 2011). In such cases, students might inadvertently reinforce incorrect language patterns without proper guidance.

Teacher feedback (TF) is the response given by instructors to help learners identify and revise mistakes and encourage them to engage in learning activities. Learners often perceive it as more valuable and reliable because teachers are always seen as subject experts (Guasch et al., 2013). In addition, TF can enhance learners' confidence in second-language writing and create a sense of encouragement and interest among students (Ruegg, 2018; Srichanyachon, 2012). However, TF also has drawbacks. Time constraints make it challenging for teachers to consistently provide meaningful feedback to all students (Gul et al., 2016; Zou et al., 2023). Besides, over-reliance on TF can hinder students' ability to critically self-assess, leading them to obediently implement corrections without analyzing their own writing (Mikume & Oyoo, 2010).

Computer-generated feedback (CF) refers to the automated responses provided by software programs to assist learners in identifying errors and suggesting improvements. Typical software programs are Grammarly (Koltovskaia, 2023), Pigai Wang (Bai & Hu, 2016), and Criterion (Li et al., 2015). CF has been found to benefit

ESL learners in several ways. Firstly, these programs provide feedback in a short time and allow students to revise and practice their writing unlimited times, thus facilitating their learning process (G. Cheng, 2017). Secondly, CF can help alleviate students' writing anxiety and embarrassment, as they receive feedback in a non-judgmental manner (Kukulska-Hulme & Viberg, 2018). Lastly, CF can guide instructors to focus on broader writing concepts rather than minor error correction, enabling them to provide more comprehensive instruction (Taskiran & Goksel, 2022). However, concerns do exist regarding CF, as it can sometimes be generic, repetitive, or even incorrect (Dikli, 2010; Jiang & Yu, 2022).

Prior studies have yielded inconsistent findings when evaluating the efficacy of TF and CF. Dikli and Bleyle (2014) asserted that TF was more concise, focused, and tailored, but CF tended to be redundant or unusable as noted in Dikli (2010). Similarly, Kaivanpanah et al. (2020) and Park (2019) discovered that TF surpassed Grammar Checker-based feedback because teachers could identify more grammatical errors and improve lexical processing. In contrast, Sistani and Tabatabaei (2023) reported that Grammarly-based feedback was outperformed due to its ability to reduce grammatical errors and even improve academic writing (Hernández Puertas, 2018). In Z. Wang and Han's (2022) study, TF improved writing quality whereas CF (i.e., Pigai Wang) could increase students' overall writing proficiency. Additionally, it was reported that TF had its unique strengths in promoting ESL learning, such as enhancing cognitive engagement (Zou et al., 2023).

However, the aforementioned studies had some issues that require further research in the following three aspects: (1) the variety of tools employed for CF across studies, such as Grammarly or Pigai Wang, may lead to conflicting results; (2) these focus predominantly on beginners and intermediate ESL learners, leaving the advanced learners unexplored; (3) the sample size is relatively small in the previous studies (e.g., 14 participants in Dikli and Bleyle [2014]), making the comparing results more unreliable. More importantly, to the best of our knowledge, no research has yet conducted a comprehensive comparison of CF, TF, and SF within a single investigation and it is still uncertain which feedback type is the most effective in terms of improving performance for ESL learners.

ChatGPT as a Computer-Generated Feedback Tool

ChatGPT is a chatbot launched by OpenAI in November 2022 (OpenAI, 2022). It adopts large language models, specifically GPT-4, to perform natural language processing tasks like writing, summarizing, translating, and answering questions (Kocoń et al., 2023; Y. Liu et al., 2023; Shen et al., 2023). ChatGPT performs well in these tasks due to its two-stage extensive training on around 45 terabytes of web data (Dwivedi et al., 2023; Zhou, Müller, et al., 2023). Beyond training by techies, ChatGPT also learns from regular users who can upvote/downvote or provide textual feedback to improve the Chatbot's responses.

Recent studies have explored the potential of ChatGPT feedback as educational assistance, focusing on both teaching and learning aspects. For teachers, ChatGPT can produce feedback relevant to improve classroom instructions, but its feedback may lack insightful and novel content (R. E. Wang & Demszky, 2023). This might be attributed to the quality of the data it was fed, as ChatGPT solely relies on statistical patterns learned from its training data (Grassini, 2023). As for students, ChatGPT feedback tends to be more detailed, fluent, and coherent, especially when evaluating data science proposal reports (Dai et al., 2023). In addition, ChatGPT feedback may improve students' task performance in other subjects such as programming problem-solving (Hellas et al., 2023) and argumentative essay writing (Su et al., 2023).

Theoretically, previous studies have suggested that ChatGPT feedback may benefit ESL learners. According to Hong (2023), ChatGPT provides instant and personalized feedback, which allows learners to make real-time improvements. Besides, S. Kim et al. (2023) claimed that feedback generated by ChatGPT is unlimited, providing students with ample opportunities for practice and refinement. In terms of language use, G. Liu and Ma (2023) found that interactions with ChatGPT can expose ESL learners to authentic language contexts, thereby enhancing their proficiency in a subconscious manner.

Empirically, however, only a limited number of studies have investigated the teachers' and students' perceptions about ChatGPT feedback within ESL learning. For instance, Mohamed (2023) and Nguyen (2023) conducted interviews with teachers, who viewed ChatGPT as an affordable and convenient tool for providing feedback. Similarly, Schmidt-Fajlik (2023) surveyed Japanese university students regarding their feelings toward ChatGPT feedback and the results showed that the majority of them expressed positive sentiments, with 89.86% of students reporting that "ChatGPT is easy to use."

Despite these findings, several issues persist in the existing literature. First, most studies on ChatGPT feedback have predominantly focused on theoretical frameworks, with few employing empirical methodologies, leading to results that are often subjective and potentially unreliable. Second, the empirical studies that do exist have primarily explored self-reported attitudes toward ChatGPT feedback, which does not adequately address the actual effectiveness of such feedback for ESL

learners. Third, the linguistic dimensions, which are essential for ESL learning, have largely been overlooked in assessing the impact of ChatGPT feedback. Given these gaps, further research is necessary to develop a comprehensive understanding of the efficacy of ChatGPT feedback, particularly in relation to linguistic dimensions, for ESL learners.

Feedback upon Written Translation

Translation is the process of transferring messages across languages and cultures. It is often regarded as the fifth basic language skill for ESL students, along with listening, speaking, reading, and writing. Improving translation quality has been a key focus of ESL learning in recent years (Drugan, 2013) and researches have shown that feedback, such as suggestions on language use, can help students improve their translations and prepare them for professional work (Alfayyadh, 2016).

Studies of translation feedback have centered on TF and SF, with few delving into CF, perhaps owing to a lack of specialized feedback systems for translation students. In terms of TF, students often reported not getting enough useful feedback from teachers (Alsahli, 2012). The insufficient teacher feedback might be owing to the labor intensity and time-consuming nature of giving feedback to a large cohort of translation students. TF requires instructors to compare the source text with the target text, which may lead to prolonged waiting periods and even demotivate students (C. Han & Lu, 2021; C. Liu & Yu, 2019). Several other studies have investigated the efficiency of SF, finding it helps student translators gain more awareness about their role in a translation task (Mellinger, 2019; Pietrzak, 2022). Nonetheless, SF is constrained by students' translation experience, making it hard for them to spot or fix mistakes (Kasperavičienė & Horbačauskienė, 2020).

In light of these challenges, the novel AI tool ChatGPT may serve as an automatic translation evaluation tool, reducing the teacher's workload and providing students with quick, detailed feedback (Frąckiewicz, 2023). ChatGPT offers real-time responses by comparing source and target texts, helping students identify mistakes and improve their self-editing skills. However, no study, to our knowledge, has directly compared TF and SF with ChatGPT feedback in terms of improving translation quality. This study is critical as it fills a gap in feedback research by examining how different feedback types influence non-native English speakers' translation quality and linguistic dimensions.

Automatic Evaluation of Translation Quality

When it comes to evaluating the translation quality, previous studies commonly relied on automatic evaluation metrics like the BLEU score (Koehn, 2010; Papineni et al., 2002). BLEU score quantifies the similarity between a candidate translation and a reference translation, with a higher score indicating closer alignment to the reference (L. Han et al., 2021). Although the BLEU score was initially designed for machine translation evaluation, it has proven applicable in assessing the quality of human-produced translation texts as well (Chung, 2020; C. Han & Lu, 2021). Even a small increase of 0.02 in the BLEU score signifies significant advancements (e.g., Bechara et al., 2011; Y. Cheng et al., 2019). Chung (2020) found a strong correlation between BLEU score and human evaluation while assessing 120 German-to-Korean translations created by 10 MTI students. Inspired by Chung (2020), C. Han and Lu (2021) further validated the feasibility of using the BLEU score to assess English-to-Chinese interpretation by students.

In addition to the BLEU score, linguistic dimensions play a crucial role in translation quality (Sofyan & Tarigan, 2019), but to the best of the researchers' knowledge, only one study has focused on this area so far. To illustrate, J. Q. Wang et al. (2021) examined the lexical performance of students' translation texts in terms of six metrics: word count, word length, lexical complexity, word range, word density, and semantic elements. However, its evaluation method did not use statistical methods (e.g., Confirmatory Factor Analysis) to prove whether these metrics could predict the lexical performance of students' translation texts. Moreover, it overlooked other linguistic features, such as syntax and cohesion. Given this, this study proposed a more comprehensive scoring system to assess the quality of student translations.

In the present study, we combined the BLEU score with three linguistic dimensions—lexicon, syntax, and cohesion—to develop a new scoring scheme for translation (Figure 1). The BLEU score is utilized to assess overall translation quality, while the linguistic dimensions are predicted using seven indicators to evaluate students' language features. For the lexicon, two indicators, word length and hypernymy for verbs are considered. Word length, as suggested by J. Q. Wang et al. (2021), serves as a measure of lexical performance, indicating that proficient translations should incorporate both longer and shorter words. Hypernymy for verbs, discussed by Ouyang et al. (2021), assesses the precision of students' interpretations. Basic texts used less specific verbs, while advanced texts employed more specific verbs, resulting in a higher average hypernymy score for verbs in the latter (Crossley et al., 2012).

Regarding syntax, three key indicators of syntactic similarity, verb phrase density, and agentless passive voice usage were identified. First, syntactic similarity can be used to reflect the fluency of translations (Polio &

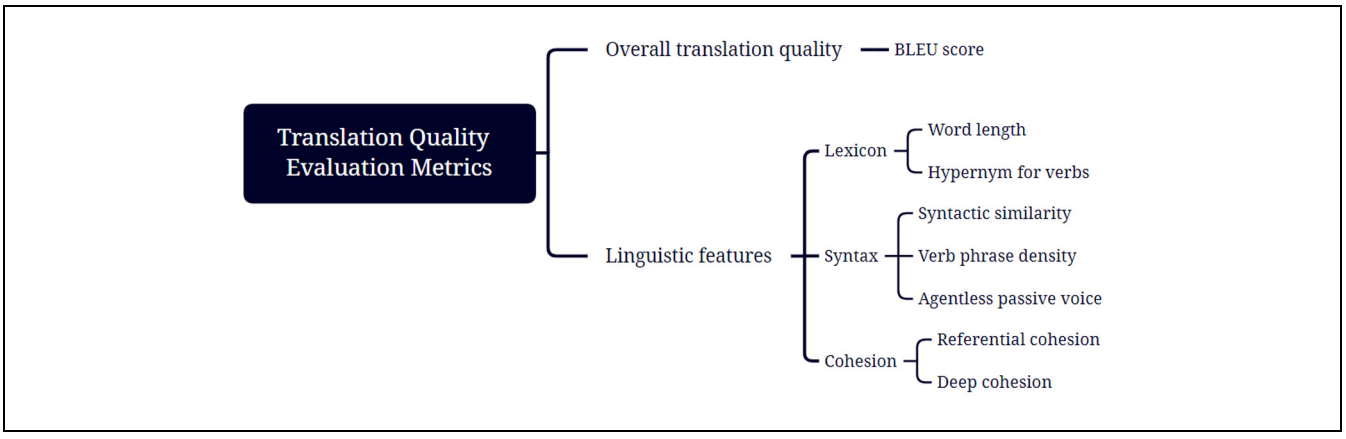

Figure 1. The new scoring scheme for translation quality.

Yoon, 2018; Sennrich, 2015). Second, verb phrase density is a significant factor to consider, as studies have shown that ESL learners tend to underutilize verb phrases in comparison to native speakers (Wu et al., 2020). Higher verb phrase density may indicate students are approaching a more native-like syntactic mastery. Third, passive voice usage was chosen, as Chinese-to-English translation often requires converting the Chinese active voice into the English passive voice (Xu et al., 2023). The capability of switching between active/passive voice in two languages shows both a strong understanding of how each language works and good translation skills.

In the domain of cohesion, two indicators were employed, namely referential cohesion and deep cohesion. Referential cohesion was chosen because it involves the use of pronouns, demonstratives, repetition, synonyms, and other cohesive devices to establish connections between ideas (Armstrong, 1991; Hall et al., 2016). Skilled translators can adapt their use of referential cohesion according to the norms of the target language to enhance clarity and coherence (Károly, 2014; Ong, 2011). Deep cohesion was included as it assesses the overall organization and connectivity of ideas by examining the causal and intentional relationships between concepts (McNamara et al., 2014). Strong deep cohesion means high logical flow and readability (Hall et al., 2016).

Research Questions

In a nutshell, the following three achievements were made in the previous studies. Firstly, CF, TF, and SF, each have unique strengths and weaknesses for improving English writing. Secondly, ChatGPT can support both ESL teaching and learning. Thirdly, theoretical studies showed that ChatGPT can deliver immediate, tailored, and interactive feedback for ESL learners. Despite these insights, it remains unknown the effectiveness of ChatGPT feedback compared with TF and SF in terms of improving students' translation quality. Hence, the present study sets out to answer the two research questions (RQ) regarding ChatGPT feedback in the context of Chinese-to-English translation:

RQ 1: Do SF, TF, and ChatGPT feedback differ in improving translation quality?
RQ 2: How do SF, TF, and ChatGPT feedback differ in improving translation lexicon, syntax, and cohesion?

Method

Participants

The present study investigated a sample of 45 MTI students (39 females and 6 males) enrolled at a prestigious university (Top 10) in China. Ranging from 23 to 26 years old ($M = 24.15$, $SD = 1.3$), all participants were native Chinese speakers who have learned English as a second language. As second-year students, they had completed 1 year of intensive translation training. Thus, they were considered intermediate student translators. Additionally, all participants passed the Test of English Majors (TEM) at band 8, the highest level for university English majors in China (Lin, 2021), which means that they were all advanced ESL learners.

Materials

The experiment included a Chinese-to-English translation task, which utilized a 424-character source text in Chinese. This text was extracted from an official press release published in 2020 on the government website of Hubei Province, China during the COVID-19 pandemic (Hubei Provincial Government, 2020). Participants were

informed that the English translation would be published alongside the Chinese source text, with the aim of providing foreign readers with updates about the pandemic. This particular document was selected as the translation material for several key reasons. First, the text difficulty was analyzed using the Chinese Resource Platform (http://120.27.70.114:8000/analysis_a), which indicated it was easily comprehensible with no major difficulties. This allowed purely testing translation capabilities of students, without confounding source text complexity. Second, as the text originated from an official press release, the language quality is high with strict editing of grammar and spelling. This prevented issues with low-quality input text from negatively impacting students' performance (Yoshimi, 2001). Third, the text has strong local relevance as it is from a Chinese provincial government website. Using regionally representative data from China provides a more accurate evaluation of the effectiveness of ChatGPT feedback in the Chinese linguistic and cultural context. In short, the selected material presented an optimal balance of difficulty, language quality, and cultural considerations to assess Chinese-to-English translation competence within the experimental constraints. Importantly, there was no existing reference translation available for this source text. This ensured students could not rely on or be influenced by official translations.

Procedure

In order to collect data, the authors collaborated with an English teacher from the aforementioned university. The experiment was conducted during a compulsory curriculum and the teacher instructed her students (the participants) to translate the provided Chinese press release into English as an assignment. Participants had experience with both SF and TF, but not with CF (ChatGPT Feedback in this context). In order to collect the data from three types of feedback (i.e., SF, TF, and ChatGPT feedback), participants were first asked to revise their initial translation texts by themselves. They were required to submit the draft translations and the revised translation texts with embedded self-feedback notes (SF-finalized version). Two weeks later, the same group of students received the notes of teacher feedback on their initial drafts and revised accordingly, generating TF-finalized versions. Finally, all these students received ChatGPT feedback (the corresponding author used ChatGPT-4 to produce feedback) on their original drafts after two weeks and produced ChatGPT feedback finalized versions. When receiving the feedback generated by ChatGPT, the author used the standardized prompt for each initial translation from students: "Please provide detailed feedback on the following student translation. Original Text: […]. Student Translation: […]." To maintain consistency across students, this same prompt was used for all draft translations, The authors, rather than the students, submitted the translation texts to ChatGPT for feedback. This approach aimed to prevent students from directly using ChatGPT, as such direct interaction could introduce numerous uncontrolled variables (e.g., variations in prompts) that might affect the results. Notably, the deliberate 2-week intervals between three submissions were strategically incorporated to avoid carry-over effects, that is, preventing recall of details in previous tasks (Bordens & Abbott, 2002).

Additionally, during the three revised processes, participants were informed not to use any AI tools (e.g., machine translation) or external resources such as dictionaries or grammar books. To reinforce compliance, students were warned that the teacher could detect the use of machine translation, which would influence their scores on this curriculum.

Data Coding

A total of 135 translation texts (45*3) were collected from three feedback revisions (ChatGPT feedback, TF, and SF). First, the data was analyzed using the BLEU score to examine overall translation quality and we followed J. Q. Wang et al.'s paradigm (2021) to calculate the BLEU score. As that BLEU score compares the similarity between the candidate translation and reference translation, we recruited four professional translators to produce four reference translations. Since BLEU automates comparison across multiple references, it allows efficient, consistent scoring of the 135 student translations in our study. Following J. Q. Wang et al.'s (2021), each student translation was scored against 4 reference translations, producing 4 individual BLEU scores per translation. We then calculated the average of these 4 scores as the final BLEU score for each translation. Averaging the scores from multiple references helped provide a robust assessment while reducing potential bias from any individual reference translation.

Following that, we utilized Coh-Metrix to obtain the data of seven linguistic indicators (i.e., word length (DESWLlt), verb density (WRDHYPv), verb and passive phrase density (DRVP, DRPVAL), syntactic similarity (SYNSTRUTt), deep cohesion (PCDCz), and referential cohesion (PCREFp) to predict three linguistic dimensions (i.e., lexicon, syntax, and cohesion; Table 1). Coh-Metrix is an automated text analysis tool (McNamara et al., 2014). According to Ouyang et al. (2021), the scores generated by Coh-Metrix were significantly correlated with human scoring of translation quality, which indicates that Coh-Metrix is reliable to collect data for testing linguistic features of translation quality.

Table 1. Coding of the New Scoring Scheme for Translation Quality.

Features	Indicators	Coding	Description
Overall quality	BLEU score	BLEU	Similarity rate between candidate and reference translations
Lexicon	Word length	DESWLlt	Average number of letters in one word
	Hypernymy for verbs	WRDHYPv	Precision degree of the verbs used
Syntax	Verb phrase density	DRVP	Incidence of verb phrases
	Agentless passive voice	DRPVAL	Incidence of agentless passive voice forms
	Sentence similarity	SYNSTRUTt	Similarity of connecting all words/phrases across sentences
Cohesion	Referential cohesion	PCREFp	Connections between sentences /clauses
	Deep cohesion	PCDCz	Causal/intentional connectives to link the whole text

Data Analysis

The study began with the use of Confirmatory Factor Analysis (CFA) to validate a model that consists of three latent factors—namely, lexicon, syntax, and cohesion. This analysis was conducted using the "*lavaan*" package (Rosseel et al., 2023) in the R (R Core Team, 2023). To assess the model fit, several statistical measures were used. The chi-square test, for example, was used to evaluate whether the covariance matrix generated by our model was consistent with the covariance matrix observed in the data. If the *p*-values were non-significant ($p > .05$), this was considered an indication of a probable good fit between the model and the data. Other measures include the standardized root mean square residual (SRMR), comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean squared error of approximation (RMSEA) indices. However, due to the small sample size ($N = 135$), the RMSEA index was treated as supplementary, because it can be unreliable in such cases (Kenny et al., 2015). Following the recommended criteria (Hu & Bentler, 1999; Humble, 2020), CFI and TLI should be greater than .90 for a good fit and greater than .95 for an excellent fit. RMSEA and SRMR should be less than .08 for an adequate fit.

After CFA, Structural Equation Modeling (SEM) was also executed using the same "*lavaan*" package in R. This was to model the causal relationships between the independent variable of "Type" (SF, TF, and ChatGPT feedback) and three latent factors (lexicon, syntax, and cohesion). The same fit indices (SRMR, RMSEA, CFI, and TLI) and criteria were employed to evaluate how well this SEM model fit the observed data.

Lastly, the study conducted two rounds of one-way analysis of variance (ANOVA) by using the EMMEANS function in the bruceR package (Bao, 2023). The first ANOVA evaluated the impact of different types of feedback (SF, TF, and ChatGPT feedback) on the three latent factors. The second ANOVA examined how these types of feedback affected the seven directly measured linguistic indicators. Conducting two separate ANOVAs enabled the study to scrutinize feedback effects at both latent and observed levels. In order to avoid type one error, we also applied Bonferroni adjustment to the alpha level (.05). If significant effects were identified in the ANOVAs, additional post-hoc Tukey HSD tests were performed to make pairwise comparisons (Lenth et al., 2023).

Table 2. Result of Structural Validity Analysis.

Features	Factor loading (λ)	*SE*	*p Value*
Lexicon			
DESWLlt	.818	.042	<.001
WRDHYPv	.619	.031	<.001
Syntax			
DRVP	.982	4.126	<.001
DRPVAL	.738	.654	<.001
SYNSTRUTt	−.180	.003	.038
Cohesion			
PCREFp	−.636	1.048	<.001
PCDCz	.696	2.101	<.001

Results

CFA Analysis

CFA results showed that the model fits the data very well, with statistical indices nearing ideal values ($\chi^2/df = 1.34$, RMSEA = .05, SRMR = .04, CFI = .99, TLI = .98). This confirmed that the three latent factors—lexicon, syntax, and cohesion—could be accurately predicted by seven observed linguistic indicators. Table 2 presents the factor loading of these variables.

SEM Analysis

Upon this validated model, SEM was applied and also showed an excellent fit to the data ($\chi^2/df = 1.21$, RMSEA = .04, SRMR = .06, CFI = .99, TLI = .98). The findings indicated that the variable of "Type" (i.e., SF, TF, and ChatGPT feedback) had a significant impact on three latent linguistic factors. As shown in

Table 2, the type of feedback is significant and positively associated with lexicon, syntax, and cohesion ($B = 0.69$, $p < .001$; $B = 0.98$, $p < .001$; $B = -1.19$, $p < .001$; Figure 2).

Evaluation of Overall Translation Quality

The results showed that the average BLEU score for students' draft translation was 0.466 while that for three revised translations based on SF, TF, and ChatGPT feedback was 0.485, 0.501, and 0.472 respectively. It is important to note that an increase of 0.02 in the BLEU score is widely considered to be a statistically significant improvement in translation quality (e.g., Bechara et al., 2011). Therefore, the revised translation based on TF scored the highest whereas those according to ChatGPT feedback scored the lowest. It indicates that TF is most effective in enhancing the overall quality of students' translations compared with ChatGPT feedback and SF.

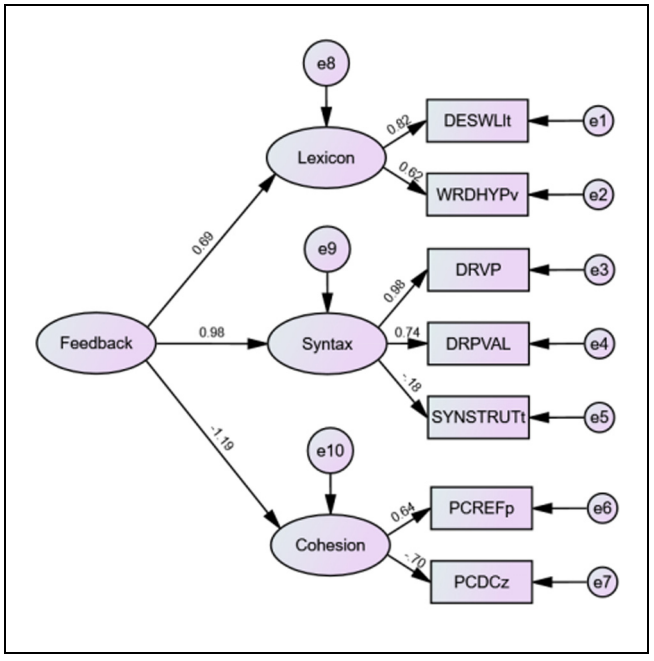

Figure 2. Structural equation model with "Type" as predictors of three linguistic factors.
Note. All parameter estimates are standardized.

Comparing Linguistic Features Across Feedback Types

Table 3 shows the result of the first one-way ANOVA. The independent variable was "Type" (SF, TF, and ChatGPT feedback) and the dependent variables were the three latent linguistic features (Lexicon, syntax, and cohesion). The results showed a significant main effect of "Type" on the lexicon ($F(2, 132) = 6.908$, $p < .01$) and syntax ($F(2, 132) = 3.173$, $p < .05$) but not on cohesion ($F(2, 132) = 2.368$, $p = .098$). As displayed in Figure 3, post-hoc Tukey tests further revealed that ChatGPT feedback yielded the highest mean score than SF and TF in the lexicon (β (ChatGPT feedback − SF) = .182, $t(132) = 2.845$, $p < .05$; β (ChatGPT feedback − TF) = .224, $t(132) = 3.494$, $p < .01$). However, no significant difference was found in the mean score between SF and TF (β (TF − SF) = −.042, $t(132) = -.648$, $p = 1.000$).

As for syntax, TF scored higher than ChatGPT feedback (β (ChatGPT feedback − TF) = −32.100, $t(132) = -2.501$, $p < .05$). TF and SF demonstrated similar performance (β (TF − SF) = 19.409, $t(132) = 1.512$, $p = .399$), while SF did not show a significant difference from ChatGPT feedback (β (ChatGPT feedback − SF) = −12.690, $t(132) = -.989$, $p = .974$).

In terms of cohesion, no significant differences were found across three feedback types (β (TF − SF) = 4.679, $t(132) = 1.941$, $p = .163$; β (ChatGPT feedback − SF) = .285, $t(132) = .118$, $p = 1.000$; β (ChatGPT feedback − TF) = −4.393, $t(132) = -1.823$, $p = .212$). This implied that ChatGPT feedback enhanced lexical capabilities more than SF and TF, while TF is optimal for developing syntactic skills compared with ChatGPT feedback.

A second round of ANOVA tests the effect of "Type" (SF, TF, and ChatGPT feedback) on seven specific observable linguistic indicators (Please see Table 4). It found that five out of the seven indicators were significantly affected by "Type": DESWLlt/word length ($F(2,$

Table 3. Three Linguistic Features Across Feedback Types.

	SF		TF		CF				
Variables	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*F(2, 132)*	*p*	η^p
Lexicon	−0.047	0.356	−0.088	0.301	0.135	0.245	6.908	<.01**	.095
Syntax	−2.240	61.014	17.17	66.111	−14.93	55.008	3.173	.045*	.046
Cohesion	−1.655	11.835	3.024	12.634	−1.369	9.614	2.368	.098	.035

Note. SF = self feedback; TF = teacher feedback; CF = ChatGPT feedback.
*$p < .05$. **$p < .01$. ***$p < .001$.

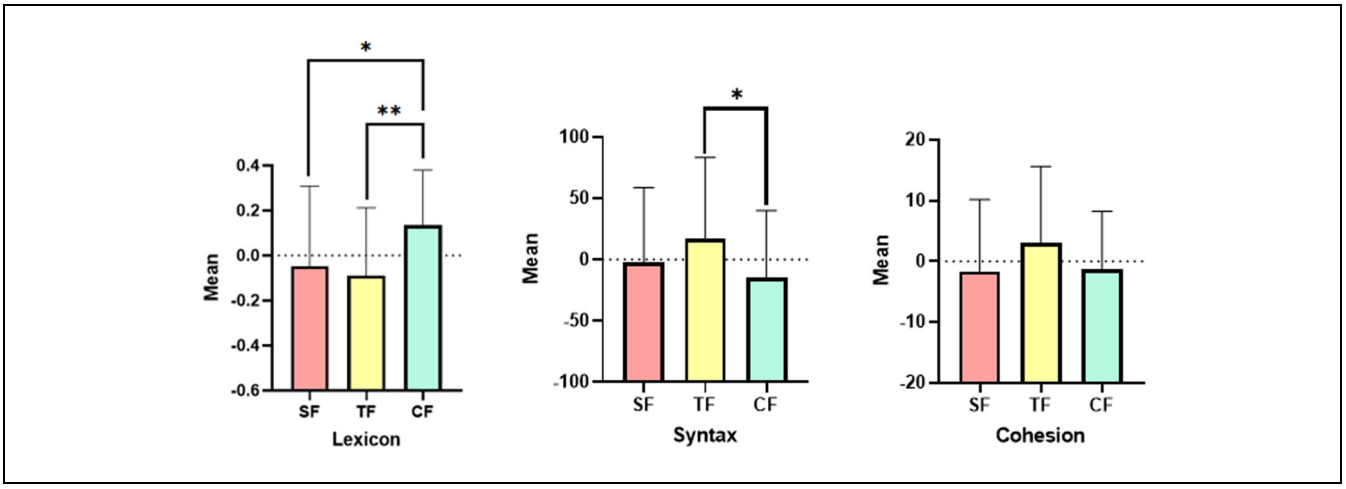

Figure 3. Mean of three linguistic features.

Table 4. Seven Linguistic Indicators Across Feedback Types.

	SF		TF		CF				
Variables	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*F(2, 132)*	*p*	η^p
DESWLlt	5.371	0.451	5.193	0.362	5.664	0.371	16.181	<.001***	.197
WRDHYPv	1.507	0.403	1.557	0.297	1.814	0.219	12.269	<.001***	.157
DRVP	149.719	62.677	171.49	67.857	137.35	57.536	3.405	.036*	.049
DRPVAL	9.594	7.635	13.274	10.51	5.627	5.928	9.686	<.001***	.128
SYNSTRUTt	0.095	0.033	0.102	0.044	0.084	0.023	2.837	.062	.041
PCREFp	63.474	27.061	61.476	25.289	76.034	23.206	4.401	.014*	.063
PCDCz	19.451	13.33	21.811	15.258	18.93	10.199	.618	.540	.009

*$p < .05$. **$p < .01$. ***$p < .001$.

132) = 16.181, $p < .001$), WRDHYPv/hypernymy for verbs ($F(2, 132) = 12.269$, $p < .001$), DRVP/verb phrase density ($F(2,132) = 3.405$, $p < .05$), DRPVAL/agentless passive voice ($F(2, 132) = 9.686$, $p < .001$) and PCREFp/referential cohesion ($F(2, 132) = 4.401$, $p < .05$). However, "Type" did not impact SYNSTRUTt/sentence similarity ($F(2, 132) = 2.837$, $p = .062$) and PCDCz/deep cohesion ($F(2, 132) = .618$, $p = .540$).

Post-hoc tests in Figure 4 show that, in the lexical level, ChatGPT feedback elicited translations with longer word length (β (ChatGPT feedback − SF) = .293, $t(132) = 3.501$, $p < .01$; β (ChatGPT feedback − TF) = .471, $t(132) = 5.634$, $p < .001$), but TF and SF did not show significant differences in word length (β (TF − SF) = −.178, $t(132) = -2.132$, $p = .104$). ChatGPT feedback also contained more specific verbs (β (ChatGPT feedback − SF) = .307, $t(132) = 4.618$, $p < .001$; β (ChatGPT feedback − TF) = .257, $t(132) = 3.860$, $p < .001$), but no significant difference was found between SF and TF (β (TF − SF) = .050, $t(132) = .758$, $p = 1.000$).

In light of syntax, ChatGPT feedback resulted in translations with less verb phrase density compared with TF (β (ChatGPT feedback − TF) = −34.140, $t(132) = -2.577$, $p < .05$), but demonstrated a similar effect as SF (β (ChatGPT feedback − SF) = −12.369, $t(132) = -.934$, $p = 1.000$). TF and SF did not show significant differences in this indicator (β (TF − SF) = 21.771, $t(132) = 1.644$, $p = .308$). Additionally, ChatGPT feedback had less agentless passive voice than the TF (β (ChatGPT feedback − TF) = −7.648, $t(132) = -4.400$, $p < .001$), but did not differ significantly with SF (β (ChatGPT feedback − SF) = −3.967, $t(132) = -2.283$, $p = .072$). TF and SF did not show a significant difference in agentless passive voice (β (TF − SF) = 3.680, $t(132) = 2.118$, $p = .108$). As for sentence similarity, ChatGPT feedback did not differ significantly from SF and TF in this dimension (β (ChatGPT feedback − SF) = −.010, $t(132) = -1.424$, $p = .470$; β (ChatGPT feedback − TF) = −.017, $t(132) = -2.366$, $p = .058$), nor was there a significant difference between TF and SF (β (TF − SF) = .007, $t(132) = .941$, $p = 1.000$).

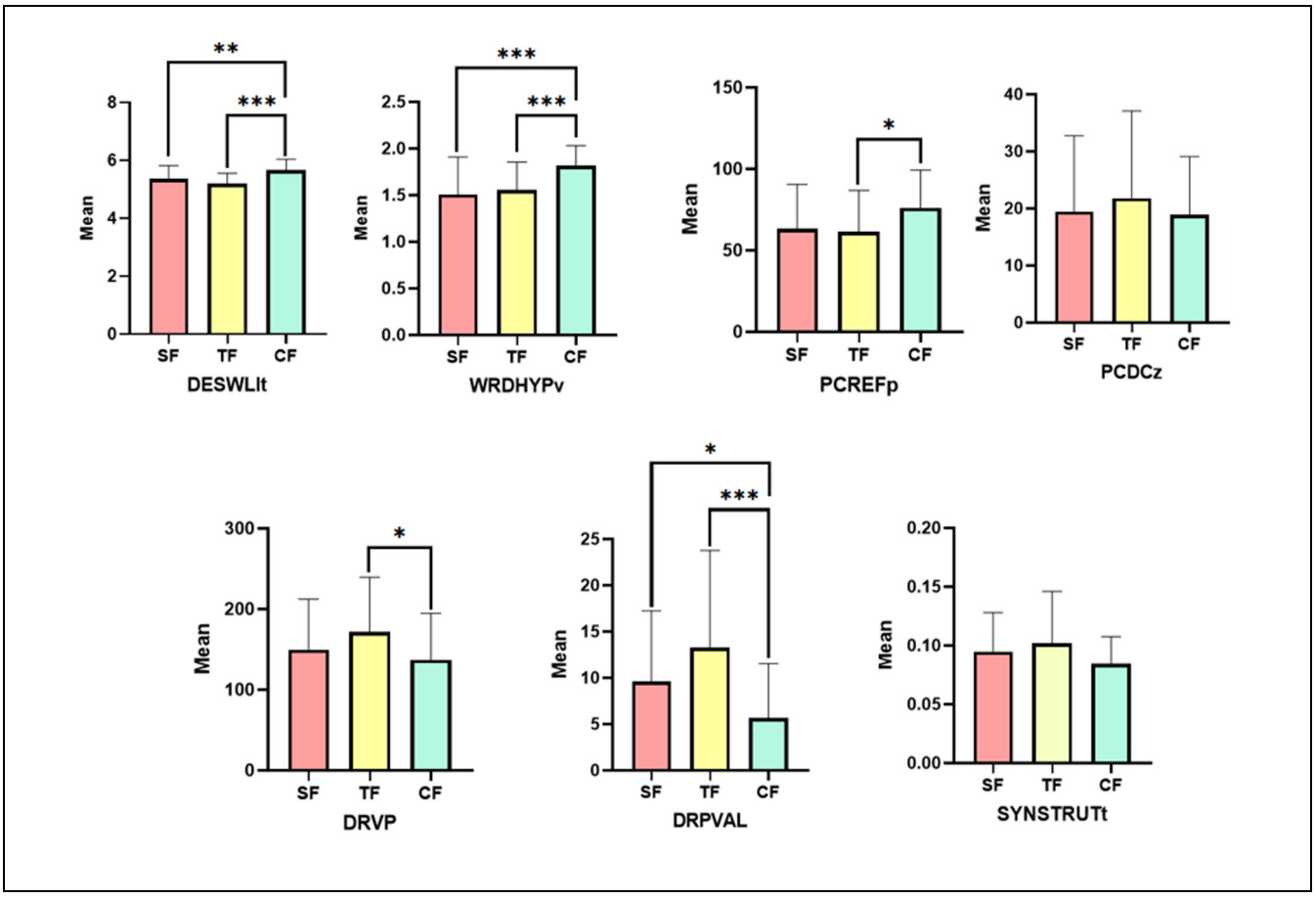

Figure 4. Mean of seven linguistic indicators.

When it comes to cohesion, ChatGPT feedback demonstrated higher referential cohesion than TF (β (ChatGPT feedback − TF) = 14.558, $t(132) = 2.736$, $p < .05$), but showed a similar effect than SF (β (ChatGPT feedback − SF) = 12.559, $t(132) = 2.361$, $p = .059$). No significant difference was found between SF and TF concerning referential cohesion (β (TF − SF) = −1.998, $t(132) = -.376$, $p = 1.000$). As for deep cohesion, no significant difference was found among SF, TF and ChatGPT feedback (β (ChatGPT feedback − SF) = −.520, $t(132) = -.188$, $p = 1.000$; β (ChatGPT feedback − TF) = −2.880, $t(132) = -1.043$, $p = .896$); (β (TF − SF) = 2.360, $t(132) = .855$, $p = 1.000$).

Discussion

The present study assessed the overall translation quality through BLEU score and relevant linguistic dimensions using Coh-Metrix, so as to evaluate ChatGPT's merits and drawbacks in generating feedback for translation practice. The results showed that both TF and SF outperformed ChatGPT feedback in improving the overall translation quality. Regarding linguistic features, we found that ChatGPT feedback showed greater gains than TF and SF in bolstering students' lexical capabilities. However, for syntactic improvement, ChatGPT was less useful than TF. Moreover, all three feedback types exhibited no significant improvements in cohesion.

We further examined the specific lexical and syntactic components that were strongly affected by each feedback type. Our findings suggested that ChatGPT feedback-guided translations exhibited greater lexical complexity, characterized by longer average word lengths and more specific verb choices compared with SF- and TF- versions. However, for syntax, TF-based translations contained denser phrasal verb patterns and increased usage of the agentless passive voice compared with ChatGPT feedback-guided versions. What follows elaborates on the above results.

Overall Translation Quality: TF > SF > ChatGPT Feedback

The results indicated that TF and SF surpassed ChatGPT feedback in improving the overall quality of student translations, as measured by the BLEU score.

This observation aligns with recent research by Bašić et al. (2023), which examined students' essay writing performance with and without the assistance of ChatGPT-3. Although our study utilized ChatGPT-4 instead of ChatGPT-3, we similarly found that ChatGPT did not enhance writing quality in either essays or translations. Furthermore, our findings are consistent with the process-oriented writing theory proposed by Hayes (2012). This theory posits that texts should undergo multiple revisions based on feedback before arriving at a final version. Such iterative revisions can foster students' reflection, critical thinking, and sense of responsibility, ultimately enhancing their overall writing abilities. In our study, the TF method involved teachers providing constructive suggestions and feedback to encourage students' reflection and critical thinking. Similarly, in the SF method, students were required to revise their work independently. In this context, ESL students clearly improved their reflection, critical thinking, and responsibility through both the TF and SF methods. In contrast, ChatGPT feedback typically offers direct responses without requiring students to engage in deeper thought. As a result, ESL students may not fully develop their writing abilities when relying solely on ChatGPT feedback.

In our study, three factors may account for the underperformance of ChatGPT feedback compared with TF and SF. First, our participants were advanced ESL learners enrolled in MTI programs. These students already possess sophisticated translation skills, making it a greater challenge for ChatGPT to provide feedback that substantially improves their translation work.

Second, ChatGPT's training data is limited by its predominantly mono-cultural, English-centric focus (Rettberg, 2022). As a result, it struggles with the nuanced demands of translation, which require not only conveying core meaning but also capturing subtle linguistic and cultural differences (Al-Sofi & Abouabdulqader, 2020; Bassnett, 2007). Our study revealed that ChatGPT frequently failed to detect errors in culturally sensitive translations. For example, it did not catch an error when students translated literally the Chinese word "*干部*" as "*cadres*." This is because the term "*cadres*" is no longer used in official Chinese documents since it has negative meanings and doesn't accurately represent modern China. A neutral and culturally appropriate translation would be "officials."

Third, we noted considerable inconsistency in ChatGPT's feedback across different student translations. While it sometimes identified issues such as incorrect verb tense or inappropriate tone, it failed to consistently highlight similar issues across multiple student translations. This inconsistency can be attributed to ChatGPT's stochastic nature, which allows it to generate different responses to the same prompt, as discussed by Jalil et al. (2023). This suggests that ChatGPT's feedback mechanism is still in a developmental stage and is not as reliable as traditional feedback methods.

Despite the aforementioned limitations, our research did identify some areas where ChatGPT exhibited strengths. For instance, it was adept at identifying redundant and verbose expressions, guiding students toward more concise and clear translations. For instance, ChatGPT spotted lengthy expressions like "*Wuhan deployed enhanced nucleic acid testing*" and suggested shortening the phrase to "*Wuhan enhanced nucleic acid testing efforts*" to avoid repetition of meaning between "*deployed*" and "*enhanced*." It also offered viable alternative expressions, drawing from its extensive vocabulary to refine unclear or non-idiomatic phrases. For instance, ChatGPT suggested students use the "*pool testing approach/method*" to replace the "*pool testing regime*" in their draft, noting that "*regime*" was not commonly used in this context in English. These observations suggest that ChatGPT has the potential to evolve into a useful automated feedback tool for language learning and translation practice.

Lexicon: ChatGPT Feedback > SF = TF

Our statistical analysis revealed that ChatGPT feedback outperformed SF and TF in improving students' lexical capability. This finding is consistent with Activity Theory (Engeström, 2001). Based on this theory, physical tools (e.g., computers) traditionally mediate human-environment interactions by facilitating physical tasks. In contrast, ChatGPT transcends this conventional role by functioning as both a mediational tool and a semiotic sign. It not only connects students with the world through technology but also provides linguistic scaffolding that directly shapes their cognitive processes. Specifically, its feedback operates symbolically—through lexical and syntactic structures—to prompt learners to expand their vocabulary repertoire and refine active language use.

In our study, one compelling reason behind ChatGPT's superior performance may lie in its extensive and diverse training data, sourced from billions of text entries such as academic articles, news reports, Wikipedia, and even literary works (Shen et al., 2023). This wide-ranging training not only equips the model with a vast lexical repertoire but also exposes it to a wide range of contextually appropriate vocabulary usage. This finding resonates with recent studies that advocated using ChatGPT for vocabulary enhancement (e.g., Baskara, 2023; Koraishi, 2023).

In fact, we found that ChatGPT feedback encouraged students to use longer words and more specific verbs. For instance, instead of employing simpler phrases like

"*separate tests*" or "*mixed tests*," it suggested more formal and contextually appropriate terms like "*individual testing*" and "*pooled testing*." It also guided students away from overly general verbs like "*deploy*" or "*ask*," steering them toward more descriptive choices like "*strengthen*" or "*request*." This guidance toward more specific and formal vocabulary likely stems from ChatGPT's exposure to complex, nuanced language during its training. This not only enriches students' active vocabulary but also improves the accuracy of their verb choices.

Conversely, both SF and TF have intrinsic limitations that make them less effective for vocabulary enhancement. For instance, SF suffers from the constraint of limited personal lexicons and less structured approaches to vocabulary building. Students often stick to the vocabulary they already know and might lack the search skills or self-discipline to incorporate new, more complex words into their translation. TF often centers on more macro-level issues, such as grammatical errors or mistranslations. Teachers may overlook refining word choices if they feel the student's translation already captures the meaning of the source text (M. Kim, 2009; Wongranu, 2017). Therefore, it may not fine-tune the vocabulary to the same degree that ChatGPT feedback does.

Apparently, no human feedback provider can match ChatGPT's data-driven vocabulary capabilities enabled by its massive training history. The evidence of marked lexical gains among students in our study strongly supports the integration of ChatGPT into translator education programs, especially for students who aim to improve their vocabulary in a nuanced and comprehensive way.

Syntax: TF = SF > ChatGPT Feedback

The result showed that TF and SF outperformed ChatGPT feedback in developing students' syntax-related skills. This finding aligns with the internal feedback model proposed by Nicol (2020), which suggests that the core process of SF involves comparing prior knowledge with external information, such as task instructions. In our study, ESL students likely synthesized their past translation experiences with the current task to refine their syntactic choices during the SF task. This effective approach explains the similar improvements in syntax observed between TF and SF.

In our observation, student translations resulting from TF and SF displayed a better grasp of complex sentence structures, such as using more sophisticated verb phrases and appropriate use of the passive voice. In contrast, translations revised via ChatGPT feedback lacked these improvements. This discrepancy can be attributed to three main factors. First of all, ChatGPT has an inherent limitation in that it cannot deeply analyze or comprehend the rules of syntax (Borji, 2023; Chomsky et al., 2023). While human instructors offer nuanced feedback based on the contextual needs of a sentence, ChatGPT's guidance tends to be more generic and superficial. For instance, it might recommend replacing one phrase with another for "better clarity," yet it frequently misses underlying syntactic issues. This was evident when we explicitly asked ChatGPT to critique the sentence structure of a complex example: "*Every community establishes special teams, creates guidelines, widely informs and engages the public, fully assumes its responsibilities, and carries out daily coordination*." Despite the sentence's obvious issues with repetition and readability, ChatGPT incorrectly praised its structure and use of parallelism.

The second limitation emerged from ChatGPT's disinclination toward passive voice. In this regard, our study aligns with AlAfnan and MohdZuki's (2023) research, revealing ChatGPT's reluctance to employ passive voice, both in its own writing and in its feedback. This indicates a more systemic limitation: if the model rarely uses passive constructions itself, it is unlikely to offer feedback that helps students understand when and how to effectively implement passive voice. However, passive voice is critical for Chinese to English translations, where Chinese sentences often lack a clear agent or subject (Hsiao et al., 2014; Zhiming, 1995). When translating to English, which often demands subjects for grammatical correctness, the ambiguity regarding the "doer" can introduce challenges. Passive voice can resolve such challenges, making translations more natural (Ke, 2023). ChatGPT falls short in this regard, unable to instruct students on how to use passive voice to tackle such challenges.

Lastly, ChatGPT lacks genre-specific feedback. The study used a news release for the translation exercise—a genre that often employs passive voice to maintain a formal, objective tone (Jacobs, 1999). In such contexts, passive constructions are not just permissible but often preferable, shifting the focus from the actor to the action or result. ChatGPT failed to offer the kind of nuanced feedback that would help students understand when and why to use passive voice in such formal settings. However, human teachers are trained to understand that different types of texts—whether news releases, academic papers, or casual conversations—have different language requirements and conventions. They understand the rationale behind these conventions and thus can impart that understanding to their students.

Cohesion: ChatGPT Feedback = TF = SF

The data demonstrated that three feedback types (ChatGPT feedback, TF, and SF) did not significantly

improve the overall cohesion in student translations. However, translations revised with ChatGPT feedback did outperform those amended with TF or SF in terms of referential cohesion. Similar to Zhou, Cao, et al.'s findings (2023), the higher use of referential cohesion indicates ChatGPT's ability to give feedback to prompt students to use more explicit linking devices between ideas, making their translations easier to follow. For instance, a translation revised with ChatGPT feedback might feature an increased frequency of synonyms or strategically employ pronouns like "*this large-scale testing*," and "*city-wide nucleic acid testing campaign*," to link sentences more clearly. This suggests that ChatGPT feedback puts considerable emphasis on enhancing referential cohesion through linguistic devices. In the case of TF and SF, the data indicated a focus on either fidelity to the source text or ensuring linguistic accuracy, rather than actively improving the text's internal coherency through referential cohesion.

Regarding deep cohesion, which involves the use of causal or intentional connectors to develop ideas, none of the feedback types exhibited significant improvement. This observation appears to conflict with Liang and Liu's (2023) findings that human translations often display better deep cohesion than machine translations. The discrepancy can be attributed to several factors. First, the scope and focus of our research are fundamentally different from those of Liang and Liu (2023). Their study directly compared final translations produced by humans and machines, whereas ours evaluated how different feedback types affected the revisions of texts initially produced by human translators.

Second, it is important to note that the technology underpinning the feedback differs between the studies. Liang and Liu (2023) relied on Google Translate for their evaluation, while we incorporated ChatGPT, a more sophisticated language model that has been shown in recent studies (e.g., Lee, 2023) to possibly surpass Google Translate in terms of translation quality.

Third, the nature of the translation task itself could be an influencing factor. Unlike free-form writing, translation is bounded by the content of the source text, which might limit the degree to which deep cohesion could be enhanced. In other words, if the source text lacks elements of deep cohesion, the translated version is the same and translators may not add more bounding words to improve cohesion. This perhaps explains why deep cohesion was not significantly improved across all samples.

Conclusion

This study compared ChatGPT feedback, teacher feedback (TF), and self-feedback (SF) for improving translation performance among advanced ESL learners. We assessed how these different types of feedback influenced overall translation quality as well as specific linguistic dimensions, including lexicon, syntax, and text cohesion. Our main findings revealed that ChatGPT feedback lagged behind both SF and TF in boosting overall translation proficiency. While ChatGPT demonstrated efficacy in some linguistic domains, such as vocabulary enrichment and referential cohesion, it was comparatively less adept in bolstering intricate syntactic competencies. The nuanced utilization of verb phrases and passive constructs, in particular, emerged as challenging areas for the AI tool.

All things considered, the findings of the current study contribute to the ongoing discussion about the role of ChatGPT in education, particularly in translator training. On a practical level, our study advocates for a blended instructional approach to translation practice. This approach combines the data-driven advantages of AI tools with nuanced, culture-aware feedback from human experts, creating a more comprehensive learning environment. By harnessing AI's efficiency alongside the insight of experienced translators, educators can provide students with a richer and more contextualized understanding of translation.

Conversely, it is essential to acknowledge potential drawbacks associated with ChatGPT in language education in translation. For instance, excessive reliance on ChatGPT may lead to a gradual decline in translators' proficiency, particularly in translating between L1 and L2. This underscores the importance of maintaining a balanced approach that combines ChatGPT with traditional translation training methods. Specifically, while ChatGPT can assist in providing quick translations and suggestions, it should not replace the critical practices of active learning, such as hands-on translation exercises, self-feedback, teacher feedback, and in-depth analysis of linguistic nuances.

Limitations and Recommendations for Future Research

This study has four primary limitations that warrant further consideration.

First, the research sample was exclusively composed of advanced ESL learners (MTI students) without controlling for specific demographic variables, and thus it is uncertain how ChatGPT feedback would perform with beginner or intermediate students, or whether demographic factors influenced the study's findings. This limitation highlights the need for future research to explore ChatGPT's effectiveness with learners of varying proficiency levels and diverse backgrounds in translation tasks.

Second, the methodology of the present study was limited to a quantitative approach. It would be more

beneficial for future studies to incorporate qualitative methods such as classroom observation, diary study, and retrospective interviews. This mixed-method approach would help gain a fuller understanding of students' perceptions, experiences, and attitudes toward different types of feedback.

Third, the assessment of overall translation quality relied solely on BLEU scores. While BLEU provides rapid and unbiased calculations, combining these scores with evaluations from human raters could enhance the reliability of the findings. Subsequent studies may consider integrating machine-generated scores with human assessments to develop more accurate methods for evaluating translation quality that reflects human judgment.

Lastly, the scope of this study was confined to a single language pair, direction, and text type. To gain a deeper understanding of ChatGPT's capabilities and limitations, future research should investigate additional language pairs, translation directions (e.g., from L2 to L1), and a broader variety of text types, including literary works. For instance, it remains uncertain whether ChatGPT would be equally effective for MTI students translating from English to Chinese. Given that ChatGPT is predominantly trained in English, the availability of training data for lower-resource languages may be limited, potentially impacting its effectiveness in those scenarios.

Acknowledgments

We also thank Linping Zhong for her constructive feedback.

Ethical Considerations

The studies involving human participants were approved by Ethics Committee of School of Foreign Languages of the participating university.

Consent to Participate

The participants provided the informed consent to participate in this study and the data were collected and mentioned in the article anonymously.

Author Contributions

SC: Conceptualization, methodology, data analysis, writing–original draft, review, and editing. TZ: Writing—review & editing, supervision

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by grants from Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX24_0351).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

The data used to support the findings of this study are available from the corresponding author upon reasonable request.

References

AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial Intelligence and Technology*, *3*(3), 85–94. https://doi.org/10.37965/jait.2023.0267

Alfayyadh, H. M. (2016). *The feedback culture in translator education: A comparative exploration of two distinct university translation programs* [Doctoral dissertation, Kent State University]. Electronic Theses & Dissertations Center. http://www.ohiolink.edu/etd/

Alsahli, F. S. (2012) *Learning and self-regulation in translation studies: The experience of students' in three contrasting undergraduate courses in Saudi Arabia* [Doctoral dissertation, The University of Edinburgh]. Moray House PhD thesis collection. http://hdl.handle.net/1842/6663

Al-Sofi, B., & Abouabdulqader, H. (2020). Bridging the gap between translation and culture: Towards a cultural dimension of translation. *International Journal of Linguistics, Literature and Culture*, *6*(1), 1–13. https://doi.org/10.21744/ijllc.v6n1.795

Armstrong, E. M. (1991). The potential of cohesion analysis in the analysis and treatment of aphasic discourse. *Clinical Linguistics & Phonetics*, *5*(1), 39–51. https://doi.org/10.3109/02699209108985501

Bai, L., & Hu, G. (2016). In the face of fallible AWE feedback: How do students respond? *Educational Psychology*, *37*(1), 67–81. https://doi.org/10.1080/01443410.2016.1223275

Bao, H. (2023, August 23). *bruceR: Broadly useful convenient and efficient R functions*. R package version 0.8.10. https://CRAN.R-project.org/package = bruceR

Bašić, Ž., Banovac, A., Kružić, I., & Jerković, I. (2023). *Better by you, better than me? ChatGPT-3 as writing assistance in students' essays*. arXiv preprint. https://doi.org/10.48550/arXiv.2302.04536

Baskara, F. R. (2023). Integrating ChatGPT into ESL writing instruction: Benefits and challenges. *International Journal of Education and Learning*, *5*(1), 44–55. https://doi.org/10.31763/ijele.v5i1.858

Bassnett, S. (2007). Culture and translation. In P. Kuhiwczak & K. Littau (Eds.), *A companion to translation studies* (pp. 13–23). Multilingual Matters.

Bechara, H., Ma, Y., & van Genabith, J. (2011). *Statistical post-editing for a statistical MT system* [Paper presentation]. *Proceedings of the 13th Machine Translation Summit*, Xiamen, China. https://aclanthology.org/2011.mtsummit-papers.35.pdf

Bordens, K. S., & Abbott, B. B. (2002). *Research design and methods: A process approach* (5th ed.). McGraw-Hill. https://psycnet.apa.org/record/2001-18329-000

Borji, A. (2023). *A categorical archive of ChatGPT failures*. arXiv preprint. https://doi.org/10.48550/arXiv.2302.03494

Cahyono, B., & Rosyida, A. (2016). Peer feedback, self correction, and writing proficiency of Indonesian ESL students. *Arab World English Journal*, *7*(1), 178–193. http://dx.doi.org/10.2139/ssrn.2804010

Cao, S., Zhou, S., Luo, Y., Wang, T., Zhou, T., & Xu, Y. (2022). A review of the ESL/EFL learners' gains from online peer feedback on English writing. *Frontiers in Psychology*, *13*, Article 1035803. https://doi.org/10.3389/fpsyg.2022.1035803

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1–6. https://doi.org/10.1080/00098650903267784

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an ESL course. *The Internet and Higher Education*, *34*, 18–27. https://doi.org/10.1016/j.iheduc.2017.04.002

Cheng, Y., Jiang, L., & Macherey, W. (2019). *Robust neural machine translation with doubly adversarial inputs*. arXiv preprint. https://doi.org/10.48550/arXiv.1906.02443

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). Noam Chomsky: The false promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Chung, H. (2020). Automatic evaluation of human translation: BLEU vs. METEOR. *Lebende Sprachen*, *65*(1), 181–205. https://doi.org/10.1515/les-2020-0009

Crossley, S. A., Allen, D., & McNamara, D. S. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, *16*(1), 89–108. https://doi.org/10.1177/1362168811423456

Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y. S., Gasevic, D., & Chen, G. (2023). *Can large language models provide feedback to students? A case study on ChatGPT*. *EdArXiv preprint*. https://doi.org/10.35542/osf.io/hcgzj

Dikli, S. (2010). The nature of automated essay scoring feedback. *Calico Journal*, *28*(1), 99–134. https://10.11139/cj.28.1.99-134

Dikli, S., & Bleyle, S. (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, *22*, 1–17. http://dx.doi.org/10.1016/j.asw.2014.03.006

Drugan, J. (2013). *Quality in professional translation* (1st ed.). Bloomsbury Publishing. https://doi.org/10.5040/9781472542014

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., & Wright, R.(2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational Al for research, practice and policy. *International Journal of Information Management*, *71*, Article 102642. https://doi.org/10.1016/i.ijinfomgt2023.102642

Engeström, Y. (2001). Expansive learning at work: Toward an activity-theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133–156. https://doi.org/10.1080/13639080020028747

Frąckiewicz, M. (2023, April 27). *The role of ChatGPT in enhancing language translation training and education*. TS2. https://ts2.space/en/the-role-of-chatgpt-in-enhancing-language-translation-training-and-education/

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7), Article 692. https://doi.org/10.3390/educsci13070692

Guasch, T., Espasa, A., Alvarez, I. M., & Kirschner, P. A. (2013). Effects of feedback on collaborative writing in an online learning environment. *Distance education*, *34*(3), 324–338. https://doi.org/10.1080/01587919.2013.835772

Gul, R. B., Tharani, A., Lakhani, A., Rizvi, N. F., & Ali, S. K. (2016). Teachers' perceptions and practices of written feedback in higher education. *World Journal of Education*, *6*(3), Article 10. http://dx.doi.org/10.5430/wje.v6n3p10

Hall, S. S., Maltby, J., Filik, R., & Paterson, K. B. (2016). Key skills for science learning: The importance of text cohesion and reading ability. *Educational Psychology*, *36*(2), 191–215. https://doi.org/10.1080/01443410.2014.926313

Han, C., & Lu, X. (2021). 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

Han, C., & Lu, X. (2021). Can automated machine translation evaluation metrics be used to assess students' interpretation in the language learning classroom? *Computer Assisted Language Learning*, 1–24. https://doi.org/10.1080/09588221.2021.1968915

Han, L., Jones, G. J., & Smeaton, A. F. (2021). *Translation quality assessment: A brief survey on manual and automatic methods*. arXiv preprint. https://doi.org/10.48550/arXiv.2105.03311

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hayes, J. R. (2012). Modeling and remodeling writing. *Written communication*, *29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hellas, A., Leinonen, J., Sarsa, S., Koutcheme, C., Kujanpää, L., & Sorva, J. (2023). *Exploring the responses of large language models to beginner programmers' help requests*. arXiv preprint. https://doi.org/10.48550/arXiv.2306.05715

Hernández Puertas, T. (2018). Teacher's feedback vs. computer-generated feedback: A focus on articles. *Language Value*, *10*(1), 67–88. http://dx.doi.org/10.6035/LanguageV.2018.10.5

Hong, W. C. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, *3*(1). Advance online publication. https://jeti.thewsu.org/index.php/cieti/article/view/103/64

Hsiao, Y., Gao, Y., & MacDonald, M. C. (2014). Agent-patient similarity affects sentence structure in language production: Evidence from subject omissions in Mandarin. *Frontiers in Psychology*, *5*, Article 1015. https://doi.org/10.3389/fpsyg.2014.01015

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hubei Provincial Government. (2020, June 2). *The 104th press conference on the prevention and control of the COVID-19 pandemic* [Conference session]. The People's Government of Hubei Province, China. http://www.hubei.gov.cn/hbfb/xwfbh/202006/t20200602_2376181.shtml

Humble, S. (2020). *Quantitative analysis of questionnaires: Techniques to explore structures and relationships* (1st ed.). Routledge. https://doi.org/10.4324/9780429400469

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Jacobs, G. (1999). Self-reference in press releases. *Journal of Pragmatics*, *31*(2), 219–242. https://doi.org/10.1016/S0378-2166(98)00077-0

Jalil, S., Rafi, S., LaToza, T. D., Moran, K., & Lam, W. (2023). ChatGPT and software testing education: Promises & perils. In *2023 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)* (pp, 4130–4137). IEEE. https://doi.org/10.1109/ICSTW58534.2023.00078

Jiang, L., & Yu, S. (2022). Appropriating automated feedback in L2 writing: Experiences of Chinese ESL student writers. *Computer Assisted Language Learning*, *35*(7), 1329–1353. https://doi.org/10.1080/09588221.2020.1799824

Kaivanpanah, S., Alavi, M., & Meschi, R. (2020). L2 writers' processing of teacher vs. computer-generated feedback. *Two Quarterly Journal of English Language Teaching and Learning University of Tabriz*, *12*(26), 175–215. https://doi.org/10.22034/elt.2020.11472.

Károly, K. (2014). Referential cohesion and news content: a case study of shifts of reference in Hungarian-English news translation. *Target. International Journal of Translation Studies*, *26*(3), 406–431. https://doi.org/10.1177/0033688210380569

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T. … Kasneci, G. (2023). ChatGPT for Good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kasperavičienė, R., & Horbačauskienė, J. (2020). Self-revision and other-revision as part of translation competence in translator training. *Journal of Language and Cultural Education*, *8*(1), 117–133. https://doi.org/10.2478/jolace-2020-0007

Ke, W. (2023). Voice conversion in CE translation. *Frontiers in Educational Research*, *6*(7), 32–36. https://doi.org/10.25236/FER.2023.060706

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kim, M. (2009). Meaning-oriented assessment of translations: SFL and its application for formative assessment. In C. V. Angelelli & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 123–157). John Benjamins. https://www.torrossa.com/en/resources/an/5016744#page=130

Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, *10*(1), 79–88. https://doi.org/10.33851/JMIS.2023.10.1.79

Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., Bielaniewicz, J., Gruza, M., Janz, A., Kanclerz, K., Kocoń, A., Koptyra, B., Mieleszczenko-Kowszewicz, W., Miłkowski, P., Oleksy, M., Piasecki, M., Radliński, L., Wojtasik, K., Woźniak, S., & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, Article 101861. https://doi.org/10.1016/j.inffus.2023.101861

Koehn, P. (2010). *Statistical machine translation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815829

Koltovskaia, S. (2023). Postsecondary L2 writing teachers' use and perceptions of Grammarly as a complement to their feedback. *ReCALL*, *35*(3), 290–304. https://doi.org/10.1017/S0958344022000179

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize ESL materials and assessment. *Language Education and Technology*, *3*(1), 55–72. http://www.langedutech.com/letjournal/index.php/let/article/view/48

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

Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, *49*(2), 207–218. https://doi.org/10.1111/bjet.12580

Lee, T. K. (2023). Artificial intelligence and posthumanist translation: ChatGPT versus the translator. *Applied Linguistics Review*. Advance online publication. https://doi.org/10.1515/applirev-2023-0122

Lenth, R., Bolker, B., Buerkner, P., Giné-Vázquez, L., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023, August 17). *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.8.8. https://cran.r-project.org/web/packages/emmeans/

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, *27*(1), 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Liang, J., & Liu, Y. (2023). Human intelligence advantages in translation competence: A corpus-based comparative study of human translation and machine translation. *Foreign Languages and Their Teaching*, (3),74-84 + 147–148. https://doi.org/10.13458/j.cnki.flatt.004946

Lin, Q. (2021). Agency fluctuations and identity transformations in Chinese English-majors on their learning trajectories. *Chinese Journal of Applied Linguistics*, *44*(4), 488–505. https://doi.org/10.1515/CJAL-2021-0031

Lipnevich, A. A., & Smith, J. K. (2022). Student-feedback interaction model: revised. *Studies in Educational Evaluation*, *75*, Article 101208. https://doi.org/10.1016/j.stueduc.2022.101208

Liu, C., & Yu, C. (2019). Understanding students' motivation in translation learning: A case study from the self-concept perspective. *Asian-Pacific Journal of Second and Foreign Language Education*, *4*(1), 1–19. https://doi.org/10.1186/s40862-019-0066-6

Liu, G., & Ma, C. (2023). Measuring ESL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, *18*(2), 125–138. https://doi.org/10.1080/17501229.2023.2240316

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). *Summary of chatgpt/gpt-4 research and perspective towards the future of large language models*. arXiv preprint. https://doi.org/10.48550/arXiv.2304.01852

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://dl.acm.org/doi/abs/10.5555/2655323

Mellinger, C. D. (2019). Metacognition and self-assessment in specialized translation education: Task awareness and metacognitive bundling. *Perspectives*, *27*(4), 604–621. https://doi.org/10.1080/0907676X.2019.1566390

Mikume, B. O., & Oyoo, S. O. (2010). Improving the practice of giving feedback on ESL learners' written compositions. *International Journal of Learning*, *17*(5), 337–353. https://doi.org/10.18848/1447-9494/CGP/v17i05/47066

Miranty, D., & Widiati, U. (2021). An automated writing evaluation (AWE) in higher education. *Pegem Journal of Education and Instruction*, *11*(4), 126–137. https://doi.org/10.47750/pegegog.11.04.12

Mohamed, A. M. (2024). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (ESL) teaching: Perceptions of EFL Faculty Members. *Education and Information Technologies*, *29*(3), 3195–3217.

Nguyen, T. T. H. (2023). ESL teachers' perspectives toward the use of ChatGPT in writing classes: A case study at Van Lang University. *International Journal of Language Instruction*, *2*(3), 1–47. https://doi.org/10.54855/ijli.23231

Nicol, D. (2020). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314

Ong, J. (2011). Investigating the use of cohesive devices by Chinese ESL learners. *The Asian ESL Journal Quarterly*, *11*(3), 42–65. https://scholarbank.nus.edu.sg/handle/10635/124356

OpenAI. (2022, November 30). *ChatGPT: Optimizing language models for dialogue*. https://openai.com/blog/chatgpt/

Ouyang, L., Lv, Q., & Liang, J. (2021). Coh-Metrix model-based automatic assessment of interpreting quality. In J. Chen & C. Han (Eds.), *Testing and assessment of interpreting: Recent developments in China* (pp. 179–200). Springer Nature Singapore. https://doi.org/10.1007/978-981-15-8554-8_9

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). *Bleu: a method for automatic evaluation of machine translation* [Paper presentation]. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, PA, USA. https://aclanthology.org/P02-1040.pdf

Park, J. (2019). An AI-based English grammar checker vs. human raters in evaluating ESL learners' writing. *Multimedia-Assisted Language Learning*, *22*(1), 112–131. https://doi.org/10.15702/mall.2019.22.1.112

Pietrzak, P. (2022). *Self-reflection as a strategy in metacognitive translator training. Metacognitive translator training: Focus on personal resources* (pp. 105–118). Springer International Publishing. https://doi.org/10.1007/978-3-030-97038-3_1

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, *28*(1), 165–188. https://doi.org/10.1111/ijal.12200

R Core Team. (2023, June 16). *The R project for statistical computing*. R Version 4.3.1. https://www.r-project.org/

Rettberg, J. (2022, December 6). *ChatGPT is multilingual but monocultural, and it's learning your values*. Jilltxt. https://jilltxt.net/right-now-chatgpt-is-multilingual-butmonocultural-but-its-learning-your-values/

Rosseel, Y., Jorgensen, T., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2023, July 19). *lavaan: Latent variable analysis*. R package version 0.6-16. https://cran.r-project.org/web/packages/lavaan/

Ruegg, R. (2018). The effect of peer and teacher feedback on changes in ESL students' writing self-efficacy. *The Language Learning Journal*, *46*(2), 87–102. https://doi.org/10.1080/09571736.2014.958190

Schmidt-Fajlik, R. (2023). ChatGPT as a grammar checker for Japanese English language learners: A comparison with Grammarly and ProWritingAid. *AsiaCALL Online Journal*, *14*(1), 105–119. https://doi.org/10.54855/acoj.231417

Sennrich, R. (2015). Modelling and optimizing on syntactic n-grams for statistical machine translation. *Transactions of the Association for Computational Linguistics*, *3*, 169–182. https://doi.org/10.1162/tacl_a_00131

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. *Radiology*, *307*(2), Article e230163. https://doi.org/10.1148/radiol.230163

Sistani, H., & Tabatabaei, O. (2023). Effects of teacher vs Grammarly feedback on Iranian ESL learners' writing skill. *International Journal of Foreign Language Teaching and Research*, *11*(45), 75–87. https://doi.org/10.30495/JFL.2023.703271

Sofyan, R., & Tarigan, B. (2019, June). *Developing a holistic model of translation quality assessment* [Paper presentation]. Eleventh Conference on Applied Linguistics (CONAPLIN 2018) (pp. 266–271). Atlantis Press. https://www.atlantis-press.com/proceedings/conaplin-18/125911470

Srichanyachon, N. (2011). A comparative study of three revision methods in ESL writing. *Journal of College Teaching & Learning*, *8*(9), 1–8. https://doi.org/10.19030/tlc.v8i9.5639

Srichanyachon, N. (2012). Teacher written feedback for L2 learners' writing development. *Silpakorn University Journal of Social Sciences, Humanities, and Arts*, *12*(1), 1–17. https://www.thaiscience.info/journals/Article/SUIJ/10850779.pdf

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, Article 100752. https://doi.org/10.1016/j.asw.2023.100752

Taskiran, A., & Goksel, N. (2022). Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal of Distance Education*, *23*(2), 120–139. https://doi.org/10.17718/tojde.1096260

Wang, J. Q., Yu, X., & Wu, W. N. (2021). Research on lexical metric-based translation quality evaluation. *Chinese Translators Journal*, *42*(05): 113–120. https://xueshu.baidu.com/usercenter/paper/show?paperid=1h260x60cy0p06h0m-r0y0r703w470986&site=xueshu_se

Wang, R. E., & Demszky, D. (2023). *Is ChatGPT a good teacher coach? Measuring zero-shot performance for scoring and providing actionable insights on classroom instruction*. arXiv preprint. https://doi.org/10.48550/arXiv.2306.03090

Wang, Z., & Han, F. (2022). The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research. *Frontiers in Psychology*, *13*, Article 909802. https://doi.org/10.3389/fpsyg.2022.909802

Wongranu, P. (2017). Errors in translation made by English major students: A study on types and causes. *Kasetsart Journal of Social Sciences*, *38*(2), 117–122. https://doi.org/10.1016/j.kjss.2016.11.003

Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, *43*, Article 100798. https://doi.org/10.1016/j.jeap.2019.100798

Xu, M., Peng, R., & Qiu, Y. (2023). A corpus-based study on Chinese translation of the be-passive in conditions of contract for EPC/Turnkey projects. *International Journal of Educational Innovation and Science*, *4*(1), 109–117. https://doi.org/10.38007/IJEIS.2023.040109

Yoshimi, T. (2001). Improvement of translation quality of English newspaper headlines by automatic pre-editing. *Machine translation*, *16*, 233–250. https://doi.org/10.1023/A:1021955327543

Yu, S., Jiang, L., & Zhou, N. (2020). Investigating what feedback practices contribute to students' writing motivation and engagement in Chinese ESL context: A large scale study. *Assessing Writing*, *44*, Article 100451. https://doi.org/10.1016/j.asw.2020.100451

Zhiming, B. (1995). Already in Singapore English. *World Englishes*, *14*(2), 181–188. https://doi.org/10.1111/j.1467-971X.1995.tb00348.x

Zhou, J., Müller, H., Holzinger, A., & Chen, F. (2023a). *Ethical ChatGPT: Concerns, challenges, and commandments*. arXiv preprint. https://doi.org/10.48550/arXiv.2305.10646

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023b). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System*, *118*, Article 103141. https://doi.org/10.1016/j.system.2023.103141

Zou, D., Xie, H., & Wang, F. L. (2023). Effects of technology enhanced peer, teacher and self-feedback on students' collaborative writing, critical thinking tendency and engagement in learning. *Journal of Computing in Higher Education*, *35*(1), 166–185. https://doi.org/10.1007/s12528-022-09337-y