FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

# A MULTIDIMENSIONAL EVALUATION OF AI-ASSISTED TRANSLATION QUALITY: COMPARING CHATGPT AND GOOGLE TRANSLATE IN ENGLISH–INDONESIAN TRANSLATION

Deri Herdawan[1*], Dhesi Wulan Sari [2], Iis Sujarwati [3]
[1,2] Politeknik Maritim Negeri Indonesia, Semarang, Indonesia
[3] Universitas Bengkulu, Bengkulu, Indonesia

deriherdawan@polimarin.ac.id

**ABSTRACT**

Limited studies have examined AI-assisted translation quality across multiple dimensions and text types in the English–Indonesian context. This study investigates the translation quality produced by ChatGPT and Google Translate in translating texts from English to Indonesian. It compares both tools across five Translation Quality Assessment (TQA) aspects: Accuracy, Equivalence, Translation Skill, Text Function, and Grammar and Style, and three text types: informative, persuasive, and instructional. Six texts were evaluated by three raters using a 19-item Likert-scale questionnaire. Quantitative analysis employed paired-samples t-tests and Cohen's d, while qualitative analysis identified translation issues. The findings show that ChatGPT outperformed Google Translate across all aspects, with statistically significant and practically meaningful differences. ChatGPT produced more natural, contextual, and communicative translations but occasionally generated additional information not found in the source text. In contrast, Google Translate tended to produce overly literal and less natural translations. The findings suggest that both tools require post-editing and are more suitable as translation aids than replacements for professional translators. This study contributes to AI-assisted translation research by providing a multidimensional evaluation of translation quality across different text types in the English–Indonesian setting.

Keywords: AI-assisted Translation, ChatGPT, Google Translate, English–Indonesian Translation

## INTRODUCTION

The advancement of technology has led to the development of Artificial Intelligence (AI) and its subsequent use in language-related activities, such as translation. AI-assisted translation tools have been widely used by academics and translators to help them resolve inter-language barriers (Amini et al., 2024; Mohamed et al., 2024; Shahmerdanova, 2025). Since its introduction, the approach used by AI-assisted translation tools has shifted from the previously rule-based approach into the current context-based approach (Shormani & Al-sohbani, 2025; Siu, 2024). As this shift happened, the use of these translation tools continues to increase and they can no longer be seen as supplementary tools, but rather as

(cc) BY

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

central instruments in communication. Such rapid integration of AI into translation activities brings up questions regarding the quality and reliability of AI-generated translations.

Currently, there are numerous AI-assisted translation tools available to use. Two of the most commonly used tools are Google Translate and ChatGPT. Google Translate is a translation tool which uses a neural machine translation system to translate texts. Its process focuses on direct transfer of meaning from the source text into the target text (Adawiyah & Wardana, 2023; Amilia & Yuwono, 2020). On the other hand, ChatGPT is designed to operate using a large language model that allows it to provide context-sensitive responses. This mechanism is capable of generating translations similar to human-like translations (Rane et al., 2023; Woodrum, 2024). Although both Google Translate and ChatGPT implement advanced AI technologies, the differing underlying mechanisms between the two translation tools may result in different translation performances. Consequently, the quality of the translations produced by these tools (such as meaning, tone, and communicative function) may be influenced.

As the popularity of AI-assisted translation tools increases, the quality of their translations remains under discussion. Previous studies have pointed out that both Google Translate and ChatGPT can produce fluent and grammatically acceptable translations, but concerns regarding the accuracy and consistency of these translations were also identified (Vries et al., 2018; Zhao, 2025). In some cases, the translation outputs may appear correct at the surface level, but they contain a few omissions, distortions, or inaccurate word choices. In addition, the capability of AI to produce translations that maintain the communicative and functional purposes needs further investigation. Therefore, there is a need for systematic and thorough evaluation of AI-generated translations. This is particularly necessary for language pairs which have distinct structural and cultural backgrounds, such as English and Indonesian.

As the next step, establishing the criteria for evaluating the quality of translation results becomes a priority. Instead of relying on intuitive judgement, a more objective Translation Quality Assessment (TQA) needs to be developed. TQA models usually propose several aspects, such as accuracy, equivalence, and grammar (Han et al., 2021; House, 2015). Initially, these models focused on linguistic comparison at the surface level, such as words, phrases, and sentences. However, more recent models have moved the focus toward multi-dimensional and functional evaluation methods. It underlines the shift in perspective in evaluating the quality of translation results. In the context of English to Indonesian translation, a comprehensive TQA model has been designed by Sofyan & Tarigan (2019). Their model proposes five aspects serving as evaluation guidelines, consisting of accuracy, meaning equivalence, translation skill, text function, as well as grammar and style. This model offers a more holistic approach as it incorporates communicative functions into the evaluation process.

Previous studies have investigated the translation performance of either ChatGPT or Google Translate. Studies on European language pairs have examined the capability of AI-assisted translation tools in specialized domains and multilingual contexts, often focusing on translation accuracy, fluency, or terminology management (Lewandowska-Tomaszczyk & Pawłowski, 2025; Sebo & de Lucia, 2024; Taira et al., 2021). While these studies provide

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

valuable insights into the strengths and limitations of AI-generated translations, their findings are largely restricted to specific text types, individual translation tools, or limited evaluation dimensions. Meanwhile, the number of studies focusing on the English-Indonesian language pair is limited. Existing studies in this context have primarily concentrated on translation errors, literal translation methods, and lexical or grammatical accuracy in Google Translate outputs (Herdawan, 2020; Napitupulu, 2017), providing relatively little attention to broader communicative and functional aspects of translation quality. In addition, the existing studies on that language pair often emphasize their analysis at the surface level accuracy without employing a multi-dimensional framework. The presence of validated instruments with confirmed validity and reliability is few in number. Therefore, there is a need to investigate the performance of both AI-assisted translation tools which apply a comprehensive TQA model within the English-Indonesian linguistic context.

In an attempt to fill the aforementioned gap, the present study compares the quality of translations produced by Google Translate and ChatGPT from English to Indonesian. To be precise, this study evaluated their performance based on five evaluation aspects as proposed by Sofyan and Tarigan (2019). Unlike previous studies that primarily focused on translation accuracy, error analysis, or single text categories, the present study adopts a multidimensional TQA framework encompassing five evaluation aspects and three different text types, namely informative, persuasive, and instructional texts. The results are expected to provide valuable empirical evidence regarding the strengths and weaknesses of both Google Translate and ChatGPT in producing translations. Theoretically, this study contributes to translation studies by extending the application of a comprehensive TQA model to the evaluation of AI-assisted translation tools within the English-Indonesian context. Practically, the findings are expected to provide translators, educators, and language practitioners with insights into the effective and critical use of AI-assisted translation technologies, particularly regarding their potential as translation support tools rather than complete substitutes for professional human translators.

## METHODS

### *Research Design and Translation Procedure*

This study implemented a comparative research design with a mixed-methods approach to investigate the quality of translations by Google Translate and ChatGPT. The source materials used in this study were in the form of documents. These documents served as the basis for evaluating and comparing the translation performance of the two AI-assisted translation systems across different text types.

Six English source texts were purposively selected to represent persuasive, instructional, and informative text types, with two texts representing each category. The texts were chosen based on four criteria: (1) they were originally written in English, (2) they represented authentic real-world texts, (3) they contained moderate linguistic complexity suitable for machine translation, and (4) they were comparable in length to facilitate balanced evaluation. The selection of multiple text types was intended to capture variations in communicative purpose and linguistic characteristics, thereby providing a more

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

comprehensive assessment of the translation performance of ChatGPT and Google Translate. While the study employed six source texts, the sample was considered sufficient for a mixed-methods design emphasizing detailed multidimensional evaluation and qualitative examination rather than large-scale corpus analysis.

*Table 1. Characteristics of source texts*

| Text Code | Text Type | Topic | Source | Word Count |
|---|---|---|---|---|
| T1 | Persuasive | Financial Literacy for Young Learners | English Academy | ~290 |
| T2 | Persuasive | Digital Literacy for Children | Alannah & Madeline Foundation | ~270 |
| T3 | Instructional | Sandwich Preparation | Crompton Kitchen Appliances Guide | ~240 |
| T4 | Instructional | Basic Vehicle Oil Change | Instructables | ~260 |
| T5 | Informative | Extensive Reading and AI-Assisted Learning | Academic Journal Article | ~310 |
| T6 | Informative | Population Ageing and Nutritional Preparation | Academic Journal Article | ~320 |

The translation process involved producing texts in Indonesian from the six English source texts. The texts were translated separately by using Google Translate and ChatGPT, resulting in two translated versions for each text. In order to maintain consistency, the texts were entered into both translation tools without any modifications. The translation results were recorded and compiled into a dataset for evaluation. In addition, these twelve translation texts were anonymized to avoid them being identified from which translation tool they were produced. This procedure was intended to minimize bias during the evaluation process.

### *Data Collection and Instrument*

The data in this study were obtained through quantitative and qualitative procedures. In order to obtain the quantitative data, a questionnaire was developed based on the TQA model developed by Sofyan and Tarigan (2019). This TQA model aims to evaluate translation quality based on five aspects, including accuracy, meaning equivalence, translation skill, text function, as well as grammar and style. The questionnaire consisted of 19 items, each representing a specific aspect (as presented in Table 2). It employed a 5-point Likert scale to measure the translations, ranging from 1 (strongly disagree) to 5 (strongly agree).

Prior to data collection, the developed questionnaire was reviewed by three experts who later also served as raters in this study. The review aimed to ensure that the questionnaire items were relevant, clear, and appropriate for assessing translation quality based on the selected TQA framework. The experts provided feedback on the wording and clarity of several items, and minor revisions were made before the instrument was finalized for use in the study.

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

*Table 2. Blueprint of the TQA questionnaire*

| Aspect | Indicator | Item Code | Number of Items |
|---|---|---|---|
| Accuracy | Meaning accuracy, completeness, absence of distortion, absence of omission/addition | A1-A4 | 4 |
| Equivalence | Lexical equivalence, phrase equivalence, sentence equivalence, terminology consistency | E1-E4 | 4 |
| Translation Skill | Sentence restructuring, handling complex structures, idiom/metaphor handling, cohesion maintenance | TS1-TS4 | 4 |
| Text Function | Communicative clarity, genre appropriateness, audience suitability | TF1-TF3 | 3 |
| Grammar and Style | Grammatical accuracy, sentence structure, naturalness, readability | GS1-GS4 | 4 |
| Total | | | 19 |

The evaluation process involved three raters with academic and professional backgrounds in translation. Two raters held a Master's degree in English Education, while one held a doctoral degree in the same field. The raters had between five and eight years of translation experience and between seven and thirteen years of university teaching experience. Their expertise in English–Indonesian translation and language education provided a strong foundation for evaluating the translation outputs generated by ChatGPT and Google Translate.

Each rater evaluated the twelve translations against the source text by employing the TQA questionnaire, independently. In addition, qualitative analysis was conducted to identify translation issues such as meaning distortion, lexical inaccuracy, grammatical errors, and shifts in communicative function. After independent coding, the raters discussed their findings to reach consensus, and representative examples were selected for reporting in the findings section.

### *Data Analysis*

Prior to data analysis, several tests were performed to ensure the reliability of the instrument and the statistical assumptions required for later computations. Table 3 summarizes those test results. Initially, the reliability of the questionnaire was evaluated by checking its internal consistency. Cronbach's Alpha analysis was conducted to determine the consistency of 19 items in measuring the construct of translation quality as defined in the TQA model. It obtained a coefficient of 0.911. According to commonly accepted reliability benchmarks (DeVellis & Thorpe, 2022), this value indicates a very good category of reliability. Therefore, the 19 items of the TQA instrument used in the present study to compare the translations between Google Translate and ChatGPT are considered statistically reliable.

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

To further assess the instrument reliability, Intraclass Correlation Coefficient (ICC) analysis was conducted to evaluate the level of agreement among the three raters (inter-rater reliability assessment). The ICC analysis was conducted using a two-way random-effects model with absolute agreement. The results showed that the ICC(2,1) value indicated moderate agreement among individual raters, while the ICC(2,k) value demonstrated good reliability when ratings were averaged across raters. These values suggest that although minor variations existed in individual judgments, the aggregated ratings provided a consistent basis for evaluating translation quality.

A normality test was conducted using the Shapiro-Wilk test. This test was selected because the number of observations obtained in this study amounted to 19. The sample size is considered small and this particular test is known for its sensitivity in detecting deviations from normality within small sample sizes (N < 50). The Shapiro-Wilk test indicated that the distribution of the difference scores did not significantly deviate from normality (p = 0.185, which exceeds 0.05).

*Table 3. Summary of preliminary test results*

| Test | Statistic | Value |
|---|---|---|
| Cronbach's Alpha | Internal consistency | 0.911 |
| ICC (2,1) | Inter-rater agreement (single rater) | 0.617 |
| ICC (2,k) | Inter-rater reliability (average raters) | 0.828 |
| Shapiro–Wilk | Normality test | p = 0.185 |

Based on these results, an inferential statistical analysis using the paired-samples t-test was done in order to determine if there are significant differences in translation quality between the two AI-assisted translation tools.

## RESULTS AND DISCUSSION

### *Overall Translation Quality Comparison*

The results of the paired-samples t-test showed that the differences in translation quality produced between ChatGPT and Google Translate were statistically significant (as presented in Table 4). In addition to being statistically significant, further analysis using the Cohen's d reported a large effect size. In other words, ChatGPT generally demonstrates better translation quality compared to Google Translate as rated by three raters, across five TQA aspects and three text types.

*Table 4. Summary of inferential statistical analysis*

| Statistic | Value |
|---|---|
| Mean ChatGPT score | 4.702 |
| Mean Google Translate score | 4.070 |
| Mean Difference | 0.632 |
| t value | 6.543 |
| p value | < 0.001 |
| Cohen's d | 1.501 |

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

One possible explanation for the substantial difference lies in the distinct technological foundations of the two systems. ChatGPT is developed using a Large Language Model (LLM) architecture which is designed to process and generate natural language. It can be tasked not only for translating, but also many other linguistic tasks. The strength of this design is that it draws on a very large corpus of texts, enabling it to capture language patterns, contexts, discourses, and semantic relations (Lyu et al., 2023; Woodrum, 2024). On the other hand, Google Translate is specifically designed for translation tasks by using a Neural Machine Translation (NMT) as its underlying mechanism. This system applies the fundamentals of mapping texts from the source language to the target language based on parallel processing, providing reliable word-by-word translation (Al-jarf, 2025; Stasimioti et al., 2020). Although current NMT systems have significantly improved translation quality compared with earlier versions, the findings of this study suggest that the broader contextual and generative capabilities of LLMs may provide advantages when translating texts that require discourse-level understanding rather than sentence-level equivalence alone.

The large effect size observed in this study also implies that the superiority of ChatGPT is meaningful from a practical perspective. A mean difference of 0.632 on a five-point scale indicates that users are likely to perceive noticeable improvements in translation quality when using ChatGPT, particularly in contexts where accuracy, naturalness, and contextual appropriateness are essential. This has implications for educational and professional settings, as translators, students, and language practitioners may benefit from using the LLM-based tools as preliminary drafting assistants while maintaining human oversight to ensure quality and cultural appropriateness.

The findings are further supported by the analysis of translation issues identified by the raters. Google Translate produced more translation issues (37) compared to ChatGPT (11). Their respective translation issues suggest that the translation results produced by these tools should not be used without proofreading. This is consistent with a previous study which underlines the necessity for professional human translators to perform a review on the AI-generated translation results (Abdelhalim et al., 2025; Ruoqi & Gochuico, 2023; Shahmerdanova, 2025). Despite their technological advancement, the findings indicate that currently ChatGPT and Google Translate cannot fully replace professional human translators. Nevertheless, while both systems can accelerate the translation process by producing rapid first drafts, ChatGPT appears to offer greater potential as a translation support because it requires fewer revisions and generates fewer quality-related issues.

### *Translation Quality by Aspects*

Table 5 presents the comparison summary between ChatGPT and Google Translate. Across the five TQA aspects, ChatGPT consistently obtained higher scores compared to Google Translate. Grammar and Style became the aspect in which the score gap was the largest between the two translation tools, with 1.000 point difference. Text Function (0.722), Translation Skill (0.708), and Equivalence (0.653) respectively came second, third, and fourth in the difference in scores. The least score difference was found in the aspect of Accuracy (0.097). The larger gaps in Grammar and Style, Text Function, and Translation Skill suggest that ChatGPT may possess stronger capabilities in producing contextually

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

appropriate and natural-sounding target texts, whereas both systems performed relatively similarly in preserving the core meaning of the source texts.

*Table 5. Score Comparison between Aspects*

| Aspect | ChatGPT (CG) | Google Translate (GT) |
|---|---|---|
| Accuracy | 4.556 | 4.458 |
| Equivalence | 4.597 | 3.944 |
| Translation Skill | 4.722 | 4.014 |
| Text Function | 4.907 | 4.185 |
| Grammar & Style | 4.778 | 3.778 |
| Overall | 4.712 | 4.076 |

Looking more closely at items within the five TQA aspects, both ChatGPT and Google Translate performed acceptably in terms of grammar, structure, and lexical equivalence of their translation results. However, most of the other items showed clearer differences between the two tools. As illustrated in Table 5, the raters reported that Google Translate produced more awkward translations that do not sound natural and fluent in the target text. Previous studies (Farahsani et al., 2021; Sutrisno, 2020) have reported similar issues in texts translated by Google Translate. The comparisons showed that better versions were frequently observed in ChatGPT-translated texts compared to their corresponding segments in Google Translate-translated texts. ChatGPT did produce similar issues, but the number was smaller. These findings further highlight that ChatGPT is currently the more effective of the two tools in performing translation tasks overall.

Despite its clear advantage over Google Translate, our analysis showed that there was one item in which ChatGPT had a negative score against Google Translate's score, indicating that the former performed worse. Within the aspect of Accuracy, the absence of additional or omitted information is rated and the three raters reported that ChatGPT produced additional information. Previous studies have found that due to its generative nature, ChatGPT may generate additional information which is not present in the source text (Alkaissi & Mcfarlane, 2023; Guerreiro et al., 2023). This phenomenon is described as "hallucination". Several examples of the translation issues identified in the present study are presented in Table 6.

*Table 6. Representative examples of translations*

| TQA Aspect | Source Text | ChatGPT | Google Translate | Preferred Translation |
|---|---|---|---|---|
| Grammar & Style | "financial literacy is no longer a luxury but a necessity" | "literasi keuangan bukan lagi sebuah kemewahan, melainkan sebuah kebutuhan" | "literasi keuangan bukan lagi kemewahan tetapi kebutuhan" | ChatGPT |
| Text Function | "schools should integrate financial literacy | "sekolah sebaiknya mengintegrasikan literasi keuangan ke | "sekolah-sekolah harus mengintegrasikan | ChatGPT |

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

| | | | | |
|---|---|---|---|---|
| | into their curriculum" | dalam kurikulum mereka" | literasi keuangan ke dalam kurikulum mereka" | |
| Translation Skill | "they often struggle to navigate the complexities of personal finance" | "mereka sering mengalami kesulitan dalam menghadapi kompleksitas keuangan pribadi" | "mereka sering kesulitan menavigasi kompleksitas keuangan pribadi" | ChatGPT |
| Equivalence | "making thoughtful financial choices rather than acting on impulse" | "membuat keputusan keuangan yang bijaksana daripada bertindak secara impulsif" | "membuat pilihan keuangan yang bijaksana daripada bertindak impulsif" | ChatGPT |
| Accuracy | Without proper guidance, children may find themselves navigating a vast sea of information without the skills needed to distinguish reliable knowledge from misleading content. | Tanpa bimbingan yang tepat, anak-anak dapat menemukan diri mereka menavigasi lautan informasi yang sangat luas tanpa memiliki keterampilan yang diperlukan untuk membedakan pengetahuan yang dapat dipercaya dari konten yang menyesatkan. Kemampuan untuk menemukan, mengevaluasi, dan menggunakan informasi melalui teknologi digital merupakan inti dari literasi digital. | Tanpa bimbingan yang tepat, anak-anak mungkin mendapati diri mereka menavigasi lautan informasi yang luas tanpa keterampilan yang dibutuhkan untuk membedakan pengetahuan yang dapat diandalkan dari konten yang menyesatkan. | Google Translate |

The translation issues identified in the present study generally can be explained due to the tools' preference of employing either literal or liberal translation approaches. Google Translate tends to prioritize literal translation strategies as it preserves the form and word arrangement from the source text. Thus, its translations are relatively close to the source text but have the weakness of sounding awkward and unnatural in Indonesian. On the other hand, ChatGPT leans toward liberal translation in performing its translation tasks as it exhibits the tendency to adjust meaning based on the context of the target language. Consequently, its translation results sound more fluent and communicative, but have the risks of generating additional information that was not explicitly present in the source text.

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

This finding strengthens the notion that literal translation supports structural fidelity, while liberal translation improves readability (Franzon, 2021; Lomaka, 2017).

The findings also show that the translation results from both AI-assisted translation tools require different kinds of post-editing efforts. In the case of ChatGPT, the issues are mainly related to additional information generated in the target text. This kind of issue can be identified by comparing the number of phrases and sentences between the paired texts. This helps determine whether information in the target text is directly traceable to the source text. Users without strong proficiency in English or Indonesian might be able to handle the post-editing in ChatGPT-translated texts. On the other hand, Google Translate produced translation issues related to inappropriate lexical choice, unnaturalness, and incorrect context-sensitive meaning. In order to identify and address these issues, high levels of proficiency in English and Indonesian are required. Therefore, it becomes necessary for human translators with good bilingual competence to handle the post-editing efforts, particularly in Google Translate-translated texts. This finding suggests that effective use of AI-assisted translation tools requires different post-editing strategies and levels of bilingual competence, depending on the characteristics of the translation issues produced by each system.

### *Translation Quality by Text Types*

Table 7 presents the comparison summary across text categories between ChatGPT and Google Translate. ChatGPT obtained higher scores compared to Google Translate in informative, persuasive, and instructional texts. The largest score difference was found in the informative text category (0.693). There was a similar score difference between the two translation tools in the persuasive text category (0.675). Meanwhile, the smallest difference was observed in the instructional text category (0.526). This is consistent with our earlier findings that ChatGPT produced better translation quality in overall.

*Table 7. Score comparison between text types*

| Text Type | ChatGPT (CG) | Google Translate (GT) |
|---|---|---|
| Informative | 4.728 | 4.035 |
| Persuasive | 4.605 | 3.930 |
| Instructional | 4.772 | 4.246 |

*Table 8. Representative Examples of Translations*

| Text Type | Source Text | ChatGPT | Google Translate | Preferred Translation |
|---|---|---|---|---|
| Instructional | “Preparing everything in advance will make the process smoother and help you avoid forgetting | “Menyiapkan semuanya terlebih dahulu akan membuat prosesnya lebih lancar dan membantu Anda menghindari lupa | “Mempersiapkan semuanya terlebih dahulu akan membuat prosesnya lebih lancar dan membantu Anda menghindari | ChatGPT |

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

| | | | | |
|---|---|---|---|---|
| | important components." | menambahkan bahan yang penting." | melupakan komponen penting." | |
| Informative | "The global population is ageing rapidly." | "Populasi dunia sedang mengalami proses penuaan dengan cepat." | "Populasi global menua dengan cepat." | ChatGPT |
| Persuasive | "For these reasons, schools should integrate financial literacy into their curriculum..." | "Oleh karena itu, sekolah sebaiknya mengintegrasikan literasi keuangan ke dalam kurikulum mereka..." | "Oleh karena itu, sekolah-sekolah harus mengintegrasikan literasi keuangan ke dalam kurikulum mereka..." | ChatGPT |

In the context of the instructional text category, the two translation systems performed relatively better compared to their performance in other text types. Instructional texts are known to emphasize procedural clarity which demands clear presentation of steps, directions, or guidance. Thus, the language tends to be direct, explicit, and systematic (Govender & Jaffer, 2021; Nickl, 2018). These language characteristics facilitate the translation process performed by both ChatGPT and Google Translate. Consequently, their translations are easily understandable, as reflected in both translation tools scoring well over 4.0. However, ChatGPT performed slightly better due to its capability of producing translations that sound more natural in the context of the target language. In contrast, Google Translate in several cases still produced expressions that sounded awkward or overly literal, particularly when dealing with certain linguistic elements such as idioms or expressions requiring contextual adjustment.

In the context of the informative text category, semantic accuracy is emphasized as the text function is to deliver information clearly and accurately to readers. Therefore, the success of its translation depends not only on lexical equivalence, but also on the capability to preserve the main information without the meaning being distorted (Ismailia, 2023; Zheng, 2017). The analysis in this study showed that ChatGPT produced better translations compared to Google Translate. In addition to producing translations that keep the original meaning, ChatGPT adjusts sentences to better match language style in Indonesian. Thus, the translations feel natural and are easy to understand. Google Translate, however, faced the same issues as it did in the previous text category. In several cases, it produced overly literal translations that did not sound natural, even though the original meaning was preserved.

In the persuasive text category, the emphasis is on pragmatic equivalence. It is essential, as the function of this kind of text is not merely to deliver information, but to influence the readers' attitude and action (Khramchenko, 2023; Kiss, 2018). Therefore, translating language elements bearing certain pragmatic nuance must be done carefully to maintain the communicative functions of persuasive text. ChatGPT's advantage became more visible in texts requiring pragmatic adjustment. An example identified in this study is the differing interpretation of the modal verb "should". In English, this modal verb represents the meaning of suggestion or recommendation, not obligation. ChatGPT

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

translated the word as "sebaiknya", which conveys a recommendation in Indonesian. On the other hand, Google Translate produced "harus" as its translation, showing a stronger degree of obligation. This highlights that ChatGPT preserves pragmatic equivalence, while Google Translate follows its way of literal translation. Thus, an inaccurate translation of a modal verb could affect the degree of persuasive effect by a text to its readers in the target language.

Overall, the findings here highlight that the unique characteristics of text type could affect how ChatGPT and Google Translate perform translation tasks. The direct and systematic language of instructional text facilitates the two tools in producing acceptable translations. In informative and persuasive text type categories, the differences between the two tools become clear as these text types demand higher capability of preserving semantic and pragmatic accuracy. The performance gap widened further when contextual interpretation was necessary. In this case, ChatGPT performed better than Google Translate. However, translation issues remain present in all translation results. This strengthens our earlier analysis that the involvement of users with strong bilingual proficiency or professional translators is necessary to ensure the accuracy, naturalness, and functional appropriateness within the context of Indonesian as the target language.

## CONCLUSION

The present study examined the comparison of translation quality produced by ChatGPT and Google Translate across five TQA aspects and three text types. Both translation tools are capable of preserving meaning from English to Indonesian. However, ChatGPT generally demonstrates better performance as it provides translations that sound natural, contextual, and communicative in Indonesian. Meanwhile, Google Translate shows a tendency to produce overly literal translations. This suggests that the different underlying mechanisms of LLM and NMT affect translation quality. It needs to be emphasized that both AI-assisted translation tools produced translation issues. Therefore, revision and post-editing efforts need to be done to improve the translations.

Based on the findings, it is suggested that ChatGPT and Google Translate be used as supporting tools rather than replacements for professional translators. The translation tools could produce a first draft quickly and efficiently while professional translators complete the task by performing revisions to ensure accuracy, naturalness, and functional appropriateness. For translation practitioners, ChatGPT may be more suitable for texts that require contextual interpretation, naturalness, and pragmatic adjustment, whereas Google Translate can still be useful for obtaining rapid literal translations that can later be refined through post-editing. In both cases, human supervision and bilingual competence remain essential to maintain translation quality and communicative effectiveness.

Despite the findings, this study has several limitations. The number of texts analyzed was limited to six texts, across three text categories, which may not fully represent the wide variety of linguistic features and communicative purposes found in real-world texts. Consequently, the findings should be interpreted cautiously and should not be generalized to all text genres or translation contexts. In addition, this study focused only on English-Indonesian language pair. Thus, it may not represent other language pairs with different linguistic and cultural contexts. Future studies should investigate a larger number of texts,

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

additional text types, other language pairs, and alternative evaluation approaches, including human post-editing analysis or user-based assessments. Such investigations would provide a more comprehensive understanding of the capabilities and limitations of AI-assisted translation tools across diverse translation settings.

**REFERENCES**

Abdelhalim, S. M., Alsahil, A. A., & Alsuhaibani, Z. A. (2025). Artificial intelligence tools and literary translation : a comparative investigation of ChatGPT and Google Translate from novice and advanced EFL student translators ' perspectives. *Cogent Arts & Humanities*, *12*(1), 1–20. https://doi.org/10.1080/23311983.2025.2508031

Adawiyah, A. R., & Wardana, L. A. (2023). Comparing Post-Editing Translations by Google NMT and Yandex NMT. Universitas Mataram. *TEKNOSASTIK*, *21*(1), 23–34.

Al-jarf, R. (2025). Google Translate Then and Now: Translations From Five Languages into English and Arabic (2012–2025). *Journal of Computer Science and Technology Studies*, *7*(12), 413–427. https://doi.org/10.32996/jcsts.2025.7.12.50

Alkaissi, H., & Mcfarlane, S. I. (2023). Artificial Hallucinations in ChatGPT : Implications in Scientific Writing. *Cureus*, *15*(2), 2–5. https://doi.org/10.7759/cureus.35179

Amilia, I. K., & Yuwono, D. E. (2020). A study of the translation of Google Translate: An error analysis of the translation of eliza riley ' s return to paradise. *Lingua Jurnal Ilmiah*, *16*(2), 1–21. https://doi.org/10.35962/lingua.v16i2.50

Amini, M., Ravindran, L., & Lee, K. (2024). Implications of Using AI in Translation Studies : Trends , Challenges , and Future Direction. *Asian Journal of Research in Education and Social Sciences*, *6*(1), 740–754. https://doi.org/10.55057/ajress.2024.6.1.67

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale Development: Theory and Applications (5th ed.)*. SAGE Publications.

Farahsani, Y., Puspita, I., Patria, R., & Jaya, H. (2021). Google Translate Accuracy in Translating Specialized Language From English to Bahasa Indonesia : A Case Study on Mechanical Engineering Terminology. *Advances in Social Science, Education and Humanities Research*, *518* (ICoSIHESS 2020), 427–435. https://doi.org/10.2991/assehr.k.210120.156

Franzon, J. (2021). The liberal mores of pop song translation: Slicing the source text relation six ways. In J. Franzon, K. A. Greenal, S. Kvam, & A. Parianou (Eds.), *Song Translation: Lyrics in Contexts* (pp. 83–121).

Govender, S. C., & Jaffer, T. (2021). Designing Instructional Text What Does Instructional Text Do? In *EdTech Books* (pp. 347–362).

Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. (2023). Hallucinations in Large Multilingual Translation Models. *Transactions of the Association for Computational Linguistics*, *11*, 1500–1517. https://doi.org/10.1162/tacl_a_00615

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

Han, L., Smeaton, A., & Jones, G. (2021). Translation quality assessment: A brief survey on manual and automatic methods. *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*, 15–33.

Herdawan, D. (2020). An Analysis on Indonesian-English Abstract Translation by Google Translate. *English Education: Jurnal Tadris Bahasa Inggris*, *13*(2), 40–53. https://doi.org/doi.org/10.24042/ee-jtbi.v13i2.7701

House, J. (2015). *Translation Quality Assessment: Past and Present* (1st ed.). Routledge.

Ismailia, T. (2023). The Analysis of Machine Translation Performance on Translating Informative Text from English into Indonesian. *Ebony - Journal of English Language Teaching, Linguistics, and Literature*, *3*(2), 129–138. https://doi.org/10.37304/ebony.v3i2.9809

Khramchenko, D. S. (2023). The power of synergy in discourse : Exploring persuasive language in English mass media. *Indonesian Journal of Applied Linguistics*, *13*(2), 368–379. https://doi.org/10.17509/ijal.v13i2.63068

Kiss, I. (2018). The Persuasive Discourse Function in the Language of Tourism. *Argumentum*, *14*, 150–162. https://argumentum.unideb.hu/2017-anyagok/kissi.pdf

Lewandowska-Tomaszczyk, B., & Pawłowski, G. (2025). Testing ChatGPT on Terminology Generation, Definitions: Translation, and Ontology Creation in German, English and Polish. *Research in Language (RiL)*, *23*(1), 320–340. https://doi.org/10.18778/1731-7533.23.20

Lomaka, A. O. (2017). Literal Approach to Translation: A Classification and Literature Review. *XLinguae*, *10*(1), 234–244. https://doi.org/10.18355/XL.2017.10.04.19

Lyu, C., Xu, J., & Wang, L. (2023). *New Trends in Machine Translation using Large Language Models : Case Examples with ChatGPT*. https://doi.org/10.48550/arXiv.2305.01181

Mohamed, Y. A., Khanan, A., Bashir, M., & Elsadig, M. A. (2024). The Impact of Artificial Intelligence on Language Translation : A Review. *IEEE Access*, *12*(February), 25553–25579. https://doi.org/10.1109/ACCESS.2024.3366802

Napitupulu, S. (2017). Analyzing Indonesian-English abstracts translation in view of translation errors by Google Translate. *International Journal of English Language and Linguistics Research Vol.5, 5*(2), 15–23.

Nickl, M. (2018). Instructional Texts: Learn How to Get Things Done. In J. Humbley, G. Budin, & C. Lauran (Eds.), *Languages for Special Purposes: An International Handbook* (pp. 322–342).

Rane, N. L., Tawde, A., Choudhary, S. P., & Rane, J. (2023). Contribution and performance of ChatGPT and other Large Language Models ( LLM ) for scientific and research advancements : a double-edged sword. *International Research Journal of Modernization in Engineering Technology and Science*, *5*(10), 875–899. https://doi.org/10.56726/IRJMETS45312

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

Ruoqi, S., & Gochuico, M. T. (2023). Translators and Their Use of ChatGPT. *Academia Lasalliana Jurnal of Education and Humanities*, *5*(1), 49–59. https://doi.org/10.55902/PLRU9197 Translators

Sebo, P., & de Lucia, S. (2024). Performance of machine translators in translating French medical research abstracts to English : A comparative study of DeepL, Google Translate, and CUBBITT. *PLoS ONE*, *19*(2), 1–13. https://doi.org/10.1371/journal.pone.0297183

Shahmerdanova, R. (2025). Artificial Intelligence in Translation : Challenges and Opportunities. *Acta Globalis Humanitatis et Linguarum*, *2*(1), 62–70. https://doi.org/10.69760/aghel.02500108

Shormani, M. Q., & Al-sohbani, Y. A. (2025). Artificial intelligence contribution to translation industry : looking back and forward. *Discover Artificial Intelligence*, *5*, 1–29. https://doi.org/10.1007/s44163-025-00487-3

Siu, S. C. (2024). Revolutionising Translation with AI: Unravelling Neural Machine Translation and Generative Pre-trained Large Language Models. In Y. Peng, H. Huang, & D. Li (Eds.), *New Advances in Translation Technology* (pp. 29–54). Springer. https://doi.org/10.1007/978-981-97-2958-6_3

Sofyan, R., & Tarigan, B. (2019). Developing a Holistic Model of Translation Quality Assessment. *Proceedings of the Eleventh Conference on Applied Linguistics (CONAPLIN 2018)*, *254*(Adv. Soc. Sci. Educ. Humanit. Res.), 266–271. https://doi.org/10.2991/conaplin-18.2019.267

Stasimioti, M., Sosoni, V., Mouratidis, D., & Kermanidis, K. (2020). Machine Translation Quality : A comparative evaluation of SMT , NMT and tailored-NMT outputs. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 441–450.

Sutrisno, A. (2020). The Accuracy and Shortcomings of Google Translate Translating English Sentences to Indonesian. *Education Quarterly Review*, *3*(4), 555–568. https://doi.org/10.31014/aior.1993.03.04.161

Taira, B. R., Kreger, V., Orue, A., & Diamond, L. C. (2021). A Pragmatic Assessment of Google Translate for Emergency Department Instructions. *J Gen Intern Med*, *36*(11), 3361–3365. https://doi.org/10.1007/s11606-021-06666-z

Vries, E. De, Schoonvelde, M., & Schumacher, G. (2018). No Longer Lost in Translation : Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications. *Political Analysis*, *26*(4), 417–430. https://doi.org/10.1017/pan.2018.26

Woodrum, C. (2024). ChatGPT and Language Translation. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in HCI*. Springer. https://doi.org/10.1007/978-3-031-60615-1_10

Zhao, S. (2025). Applications and Challenges of ChatGPT in Translation: A Multidimensional Exploration of Theory and Practice. *Sino-US English Teaching*, *22*(2), 35–41.

FOSTER JELT
Journal of English Language Teaching

Copyright © The Author(s)
Vol. 7, No. 4, October 2026
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

https://doi.org/10.17265/1539-8072/2025.02.01

Zheng, W. (2017). Translation Strategies for Texts of Science and Technology. *International Conference on Humanities Science, Management and Education Technology (HSMET 2017)*, *96*(HSMET), 32–36. https://doi.org/10.2991/hsmet-17.2017.6