ARGOPURO

THE EFFECTIVENESS OF USING DEEPL AS A TRANSLATION TOOL ON THE ACCURACY OF ENGLISH TRANSLATIONS AMONG ENGLISH DEPARTMENT STUDENTS AT UNIVERSITAS BHINNEKA PGRI

Shafna Pratiwi Junaidi, Ayu Rizki Septiana
Universitas Bhinneka PGRI[1,2]
shfn.jun@gmail.com[1] , ayurizki.septiana@gmail.com[2]

Abstract

This study aims to examine the effectiveness of DeepL Translator as a tool for improving the accuracy of English translation among sixth-semester students of the English Department at Universitas Bhinneka PGRI. Utilizing a pre-experimental one-group pre-test–post-test design, this quantitative research involved 16 participants. The research instruments consisted of pre-test and post-test translation tasks designed to assess students' translation accuracy before and after using DeepL Translator. These tasks were evaluated by two independent raters using a detailed scoring rubric.

Data analysis included descriptive statistics, normality testing using the Shapiro-Wilk test, and hypothesis testing using the Wilcoxon Signed-Rank Test, as the pre-test scores did not meet the assumption of normality. The results indicated a significant improvement, with the average score increasing from 59.31 (pre-test) to 83.44 (post-test). This difference was statistically significant, with a Wilcoxon significance value of .000 ($p < 0.05$). Furthermore, inter-rater reliability was high, as shown by a Cronbach's Alpha of 0.877 and an Intraclass Correlation Coefficient (ICC) of 0.833, indicating strong consistency between raters.

Keywords: Translation tools, DeepL Translator, Translation Accuracy

INTRODUCTION

In today's globalized world, the rapid advancement of technology—especially in the field of Artificial Intelligence (AI), has revolutionized many aspects of education, including foreign language learning. English, as the most widely spoken international language, plays a critical role in academic and professional domains, prompting learners to develop not only grammatical proficiency but also applied language skills such as translation (Council, 2013; Fitria, 2021). The integration of AI tools in language education has opened new avenues for improving translation practices, offering students support in bridging linguistic and cultural gaps (Syarifudin, 2023).

Translation remains one of the most essential competencies in foreign language acquisition, particularly in academic contexts where students must accurately convey meaning between source and target languages. However, this task poses considerable challenges. Translating texts is not merely a linguistic exercise but involves conveying meaning, tone, and nuance naturally and fluently in the target language (Septiana, 2020). Research has shown that students often face difficulties in achieving accuracy, particularly when translating complex or

academic content, necessitating the need for tools that can enhance their translation skills (Munawwarah & Martriwati, 2024).

To address these challenges, machine translation (MT) tools such as Google Translate and DeepL have emerged as valuable aids. While Google Translate remains one of the most widely used platforms, DeepL, launched in 2017, has gained recognition for its superior translation accuracy and fluency, largely due to its use of Neural Machine Translation (NMT) technology (Nugraheni & Sutrisno, 2024). DeepL employs deep learning algorithms that allow it to capture contextual nuances and generate more natural-sounding output. Compared to traditional MT systems, NMT enhances syntactic coherence, word choice, and grammatical agreement, making it suitable for academic translation tasks (Wu et al., 2016; Yang et al., 2020).

Several studies have examined the use of DeepL in educational contexts. Sidiq and Syafryadin (2024) explored students' perceptions and found that while DeepL helped improve vocabulary and simplified the translation process, many students remained skeptical about the accuracy of its output. This perception was echoed in the work of Asmara and Kembaren (2024),who observed that students preferred DeepL for translating academic texts due to its higher accuracy compared to other tools, yet they still relied on their knowledge to review and revise the translations. These findings highlight a research gap: while prior studies focused on perceptions, few have provided empirical evidence assessing the measurable effectiveness of DeepL in improving translation accuracy.

Moreover, DeepL's ability to provide context-sensitive alternatives, suggest appropriate diction, and support various document formats (e.g., PDF, DOCX) adds to its appeal in academic settings (Kirana, 2024). The tool's integration of AI and contextual analysis is especially useful for English Department students, who are often required to translate theses, articles, or technical content with a high degree of accuracy. Nonetheless, questions remain regarding the actual impact of such tools on students' performance when evaluated objectively.

To address this gap, the present study investigates the effectiveness of using DeepL as a translation tool in enhancing the accuracy of English translations among sixth-semester students at Universitas Bhinneka PGRI. By employing a one-group pre-test–post-test design, this research seeks to move beyond subjective perceptions and offer concrete, quantitative data on the tool's impact. This study contributes to the growing body of knowledge on AI-supported language learning and aims to provide both students and lecturers with evidence-based recommendations for integrating DeepL in translation instruction.

METHODOLOGY
Research Design
This study employed a pre-experimental one-group pre-test–post-test design. This method was chosen to examine the effectiveness of DeepL Translator by comparing students' translation accuracy before and after treatment (Creswell, 2014).
Variables
  a. Independent Variable (X): Use of DeepL Translator
  b. Dependent Variable (Y): Accuracy of English translations
Population and Sample
  a. Population: Sixth-semester students of the English Education Department at Universitas Bhinneka PGRI in the academic year 2024/2025
  b. Sample: 16 students
  c. Sampling Technique: Total sampling (Cohen et al., 2007)
Procedure of Data Collection
  a. Pre-Test: Students translated 10 sentences and 1 paragraph without using any translation tools.
  b. Treatment: Students received instruction and practice using DeepL Translator in three 90-minute sessions.
  c. Post-Test: Students completed a test identical to the pre-test.

Table 1. Teaching Plan

Meeting	Activity	Details
1	Introduction to translation and tools	– Explain types and methods of translation- Demonstrate DeepL Translator usage
2	Hands-on practice	– Students translate academic texts with and without DeepL- Analyze translation results
3	Evaluation & reflection	– Assess translation using scoring rubric- Reflect on DeepL usage and translation performance

Table 2. Scoring Rubric of the Instrument

Criteria	Maximum Score
Meaning Equivalence	25
Grammar Accuracy	25
Diction Choice	25
Naturalness / Fluency	25
Total	100

(*Adapted from* Samir & Tabatabaee-Yazdi, 2020)

Data Analysis Techniques
a. Descriptive Statistics: To analyze mean scores before and after treatment.
b. Normality Test: Shapiro-Wilk Test.
c. Hypothesis Testing: Wilcoxon Signed-Rank Test due to non-normal distribution of pre-test scores.
d. Reliability Testing: Cronbach's Alpha ($\alpha = 0.877$) and Intraclass Correlation Coefficient (ICC = 0.833) confirmed inter-rater reliability (Gwet, 2014).

RESULTS

This section presents the findings of the study regarding the effectiveness of DeepL Translator on students' English translation accuracy. The data were obtained from pre-tests and post-tests administered to 16 sixth-semester students of the English Education Department at Universitas Bhinneka PGRI.

1. Descriptive Statistics of Pre-test and Post-test Scores

To evaluate the students' translation performance, both pre-test and post-test scores were analyzed. The mean score increased notably after treatment.

Table 3. Descriptive Analysis

Test	N	Minimum	Maximum	Mean	Std. Deviation
Pre-test	16	50.5	63.0	59.313	3.7766
Post-test	16	78.5	88.0	83.437	3.0325

The results show a clear improvement in students' translation performance after using DeepL.

2. Reliability Test

To ensure consistency between the two raters who assessed students' translations, reliability testing was conducted using Cronbach's Alpha and Intraclass Correlation Coefficient (ICC).

Table 4. Reliability Test

Cronbach's Alpha	N of Items
0.877	2

Table 5. Intraclass Correlation Coefficient

ICC Type	Value	95% CI (Lower–Upper)	Sig.
Single Measures	0.791	0.496 – 0.922	.000
Average Measures	0.833	0.664 – 0.959	.000

These values indicate excellent inter-rater reliability (Gwet, 2014).

3. Normality Test

The Shapiro-Wilk Test was used to assess the normality of the data. The pre-test data did not meet the assumption of normality.

Table 6. Shapiro-Wilk Test of Normality

Test	Statistic	df	Sig.
Pre-test	0.855	16	0.016
Post-test	0.942	16	0.380

Since the pre-test data is not normally distributed ($p < 0.05$), a non-parametric test was applied for hypothesis testing.

4. Hypothesis Testing (Wilcoxon Signed-Rank Test)

To examine whether there was a significant difference between the pre-test and post-test scores, the Wilcoxon Signed-Rank Test was conducted.

Table 7. Wilcoxon Ranks

Comparison	N	Mean Rank	Sum of Ranks
Negative Ranks	0	-	-
Positive Ranks	16	8.50	136.00
Ties	0	-	-

Table 8. Wilcoxon Test Statistics

Z	Asymp. Sig. (2-tailed)
-3.519	0.000

The significance value (0.000) is below 0.05, indicating a statistically significant difference in students' translation accuracy before and after using DeepL Translator.

The results of this study demonstrate a statistically significant improvement in students' English translation accuracy after using DeepL Translator. The mean score increased from 59.31 in the pre-test to 83.44 in the post-test, as confirmed by the Wilcoxon Signed-Rank Test (p = 0.000). This improvement suggests that DeepL had a meaningful effect on enhancing the students' translation performance.

These findings align with prior studies highlighting the benefits of Neural Machine Translation (NMT) tools in language learning. Nugraheni and Sutrisno (2024) emphasize that DeepL's NMT system allows it to generate more natural and contextually appropriate translations compared to other tools. The students in this study were able to improve their grammar, vocabulary usage, and fluency, which suggests that the AI-enhanced output supported their understanding of language patterns and word choices.

Additionally, the results are consistent with the research of Asmara and Kembaren (2024), who reported that students found DeepL more accurate and effective than Google Translate when working with complex academic texts. In the present study, students were trained to analyze and revise DeepL's output using a rubric, which likely contributed to their improved

scores. As Bowker and Ciro (2019) argue, machine translation tools can be pedagogically effective when used in conjunction with critical evaluation and guided reflection.

Moreover, the students' consistent post-test improvement across all measured aspects—meaning equivalence, grammar, diction, and fluency indicates that DeepL not only facilitated translation but also contributed to skill development. Payung and Sukarno (2025) support this perspective, suggesting that DeepL helps students enhance their writing and editing strategies by offering contextual alternatives and feedback.

However, this study also acknowledges that while DeepL improved translation accuracy, it may not replace human translation competence entirely. As found by Sidiq and Syafryadin (2024), students still felt the need to verify grammar and meaning manually, indicating that full reliance on the tool may not always yield optimal results. This reinforces the need for instructional support and critical engagement when integrating AI-based tools into language education.

In short, the use of DeepL Translator, when combined with structured learning activities, rubric-based evaluation, and teacher guidance, can be an effective method for improving students' translation performance. It empowers students not only to produce accurate translations but also to become more aware of linguistic features and translation strategies.

CONCLUSION

This study concludes that the use of DeepL Translator significantly improves the accuracy of English translations among sixth-semester students of the English Education Department at Universitas Bhinneka PGRI. The increase in post-test scores, supported by the Wilcoxon Signed-Rank Test (p = 0.000), confirms the effectiveness of DeepL in enhancing students' performance in aspects such as meaning equivalence, grammar, diction, and fluency. This outcome suggests that DeepL, as a Neural Machine Translation tool, provides meaningful support for students' academic translation tasks and can serve as a beneficial aid in language learning when integrated with structured instruction and critical evaluation practices.

Based on the results, it is suggested that lecturers encourage students to use DeepL Translator as a supplementary tool for translation assignments while also providing constructive feedback to ensure students maintain awareness of grammatical accuracy and context. Furthermore, future researchers are recommended to explore the long-term impact of AI-based translation tools like DeepL in various educational settings, and to develop strategies for integrating these technologies effectively into language curricula to enhance students' critical thinking and language competence

BIBLIOGRAPHIE

Asmara, D. S. M., & Kembaren, F. R. B. (2024). Student's Perception Towards the Use of DeepL Translator in Writing Thesis or Journal for English Education Students. *Ijlecr - International Journal of Language Education and Culture Review*, *10*(1), 117–126. https://doi.org/10.21009/ijlecr.v10i1.47937

Bowker, L., & Ciro, B. J. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Emerald Publishing. https://doi.org/https://doi.org/10.1108/978-1-78756-721-420191004

Cohen, L., Manion, L., & Morrison, K. (2007). Research Methods in Education. In *Research Methods in Physical Activity and Health*. https://doi.org/10.4324/9781315158501-17

Council, B. (2013). The English Effect: The impact of English, what it's worth to the UK and why it matters to the world. In *British Council*. https://doi.org/10.2307/jj.8085283.10

Creswell, J. W. (2014). *Research Design (Qualitative, Quantitative & Mixed Methods Approaches) 4th Edition*.

Fitria, T. N. (2021). The Use Technology Based on Artificial Intelligence in English Teaching and Learning. *ELT Echo : The Journal of English Language Teaching in Foreign Language*

Context, 6(2), 213–223. https://doi.org/10.24235/eltecho.v6i2.9299

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability Fourth Edition* (Fourth Edi).

Kirana, A. (2024). *The Impact of Using DeepL Translator On EFL Students' Writing*. http://repository.uinsu.ac.id/id/eprint/23162

Munawwarah, M., & Martriwati. (2024). *Students' Perception of Using DeepL as Machine Translation in English Learning*. *5*(2), 284–295. https://doi.org/10.22236/ellter.v5i2.15813.g4779

Nugraheni, Y. M., & Sutrisno, A. (2024). DeepL vs ChatGPT : Machine Translation Evaluation. *Prologue: Journal on Language and Literature*, *10*(2), 411–426. https://doi.org/https://doi.org/10.36277/jurnalprologue.v10i2.174

Payung, N. F., & Sukarno. (2025). *Enhancing EFL Students ' Writing Skills : The Impact of DeepL Translation*. *13*(1).

Samir, A., & Tabatabaee-Yazdi, M. (2020). Translation Quality assessment rubric: A Rasch Model-based Validation. *International Journal of Language Testing*, *10*(2), 101–128.

Septiana, A. R. (2020). Exploring Students' Strategies and Methods in Translating Indonesia-English Text. *Journal of English Language Teaching Learning and Literature LinguA-LiterA*, *3*(2), 62–73. https://jurnal.stkippgritrenggalek.ac.id/index.php/lingua/article/view/34

Sidiq, F. A., & Syafryadin. (2024). Students' perception of using DeepL for translating English text. *Eltin Journal: Journal of English Language Teaching in Indonesia*, *12*(1), 139–148. http://www.deepl.com

Syarifudin, A. S. (2023). *Challenges and Opportunities of Implementing AI in Language Learning in Indonesia*. *3*(1). https://doi.org/10.4018/979-8-3693-3989-3.ch004

Wu, Y., Schuster, M., Chen, Z., Le, Q. V, Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., … Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. 1–23. http://arxiv.org/abs/1609.08144

Yang, S., Wang, Y., & Chu, X. (2020). *A Survey of Deep Learning Techniques for Neural Machine Translation*. http://arxiv.org/abs/2002.07526