مـن مـوضـوعـات الـعـدد

- **Investigating the Pedagogical Effectiveness of AI-Based Translation Tools in Fostering EFL Writing Skills and EFL Vocabulary Skills at Al Madinah International University**

- تأثير التعلم العملي في تحسين مهارات الطلاب في التحدث باللغة العربية: الجامعة الإسلامية العالمية بماليزيا أنموذجا

- استخدام الألعاب اللغوية في تنمية مهارة القراءة لدى طلاب المرحلة الابتدائية من وجهة نظر المعلمين في كوالالمبور: دراسة وصفية تحليلية

- الأمنيات المستحيلة في القرآن الكريم دراسة أسلوبية

Investigating the Pedagogical Effectiveness of AI-Based Translation Tools in Fostering EFL Writing Skills and EFL Vocabulary Skills at Al Madinah International University

Gellan Gamal Atwa[1*]
[1] Lecturer, Al-Madinah International University
[*] gellan.abdel@mediu.my

Abstract: Artificial intelligence (AI) tools are increasingly converting language education by providing modern tools that support language learning and writing development. Among these tools, AI-driven translation tools have gained attention for their potential to assist English as a Foreign Language (EFL) learners in improving writing performance and expanding vocabulary knowledge. Despite their growing use, experimental evidence on their efficacy in higher education contexts remains limited. This study investigated the impact of AI-driven translation tools on EFL students' writing skills and vocabulary acquisition at Al Madinah International University. Specifically, it aimed to examine whether the use of AI translation tools—DeepL, ChatGPT, and Google Gemini—could significantly improve students' writing performance, vocabulary development, and learning autonomy. A quasi-experimental design was employed with 30 EFL students divided into a control group (N = 15) and an experimental group (N = 15). Data were collected through pre- and post-intervention writing and vocabulary assessments and analyzed using the Mann–Whitney and Wilcoxon Signed-Rank tests. The findings revealed statistically significant improvements in both writing and vocabulary outcomes among students who used AI translation tools. The intervention produced notable gains in overall writing performance, particularly in the "Content and Ideas" sub-skill (effect size = 0.750). Similarly, the treatment demonstrated a substantial impact on vocabulary learning (effect size = 1), improving most sub-skills except "Discriminating Vocabulary Nuances." Participants also reported increased confidence and autonomy in their learning process. These results suggest that AI-driven translation tools can serve as effective supplementary resources in EFL instruction, with important implications for curriculum design and the strategic integration of AI technologies in higher education language learning environments.

Introduction

The teaching and learning of English in contexts where it is not the predominant language of communication is known as English as a Foreign Language (EFL). EFL is taught in environments where students are not exposed to English outside the classroom, unlike ESL, which is frequently taught in English-speaking nations. EFL has become an essential part of the curriculum in schools around the world as the need for English proficiency rises with globalization. It is crucial for cross-cultural communication, professional growth, and academic achievement. The four core language skills—speaking, listening, reading, and writing—as well as fostering in students a sense of cultural awareness and motivation, are the main goals of practical EFL training.

The primary objective of the current study is to enhance EFL writing skills and vocabulary acquisition. Wu and Halim (2024) assert that EFL writing is a multifaceted process influenced by various factors, including individual characteristics (cognition, affect, and working memory) and social environmental factors (learning and curricular environments). It is vital to include cognitive and emotional factors in the study of EFL writing performance, such as anxiety, joy, and boredom. Emotional learning also affects writing performance, along with cognitive factors.

A substantial correlation exists between vocabulary comprehension and writing proficiency in English as a Foreign Language (EFL) research. The breadth and depth of vocabulary directly influence the quality, coherence, and effectiveness with which EFL learners articulate concepts. Writing assignments, especially those focused on vocabulary development, can significantly enhance vocabulary acquisition and overall linguistic proficiency. Lexicon as a Foundation: A sophisticated and expansive vocabulary is crucial for effective writing, enabling students to produce texts that are more intricate, accurate, and coherent. Forecasting Ability: A robust vocabulary is significantly associated with writing proficiency, particularly in the structural and linguistic aspects of composition. Nevertheless, its impact may be indirect, often mediated by reading comprehension

skills (Dhuli, Lamo, & Larsari, 2023; Balle & Olsen, 2024).

Incorporating writing tasks that focus on vocabulary and promoting note-taking can significantly enhance vocabulary and writing abilities in EFL students. Educators need to recognise that although writing strategies enhance engagement, they might not directly lead to immediate vocabulary improvements unless paired with more thorough processing methods. Thus, vocabulary comprehension and writing abilities are strongly connected in EFL education. Engaging learners with vocabulary through writing activities enhances writing skills and promotes better vocabulary learning (Tu & Jiang, 2024; Kakihana, 2024; Jin & Webb, 2021; Jurado, 2023; Balle & Olsen, 2024 & Albelihi, 2022).

Dhuli, Lamo, and Larsari (2023) investigated the importance of vocabulary comprehension in enhancing the writing abilities of ESL/EFL learners. It examines theoretical viewpoints and outlines empirical research that shows a distinct relationship between vocabulary size and writing performance. The results show that students with a larger vocabulary excel at writing tasks, whereas those with a restricted vocabulary struggle. The paper wraps up by highlighting the importance of vocabulary enhancement techniques for both students and teachers to boost writing abilities.

Translation-based approaches are commonly employed in EFL (English as a Foreign Language) settings to improve writing and vocabulary enhancement. Studies consistently indicate that translation techniques—whether conventional or digital—can improve EFL students' writing abilities and vocabulary development, particularly when paired with feedback and effective instructional methods. Some of these strategies focus on the effectiveness of EFL writing skills and assert that Translation tasks incorporating machine translation (MT) and resources such as DeepL have proven to significantly enhance EFL learners' writing skills, minimise grammatical mistakes, and improve revision techniques. These methods also enhance students' confidence and motivation in writing, encouraging them to be more independent and involved in the writing process. Translation techniques help students perceive writing as a process, encouraging multiple drafts and edits, resulting in improved writing. Even proficient EFL learners utilise translation as a support mechanism, frequently converting

concepts from their native language (L1) into English (L2) to better organise and convey their ideas (Payung & Sukarno, 2025; Zhang, 2023; Lee,2021& Payung & Sukarno, 2025).

Literature review

Translation-based approaches utilise various educational theories to shape the ways translation is learned and taught. These frameworks strive to connect translation abilities with broader language skills, meet the needs of multilingual students, and adapt to changing educational settings, such as digital and mobile learning environments. Essential Theoretical Frameworks in Translation Education Constructivist Paradigm: Highlights the importance of student involvement and learning via problem-solving, framing translation as a process in which learners develop knowledge by tackling authentic translation issues. Problem-Based Learning: Promotes student engagement with real translation issues, enhancing both translation skills and language proficiency through comprehensive and focused teaching approaches (Jiménez, 2019). Translation-based pedagogical approaches are grounded in diverse theoretical frameworks, including constructivism, activity theory, translingual pedagogy, threshold concepts, and psycholinguistics. These frameworks collectively support active, context-sensitive, and inclusive translation learning, empowering both students and teachers to navigate complex linguistic and cultural landscapes (Martín et al., 2019).

Artificial intelligence translation methodologies have significantly progressed, characterized by two predominant paradigms: statistical machine translation (SMT) and neural machine translation (NMT). Statistical Machine Translation (SMT) employs statistical models derived from extensive bilingual corpora, including word-based, phrase-based, and syntax-based models, to forecast translations through probability distributions (Mondal, Zhang, Kabir & Dai, 2023 & Wei, 2024). NMT employs deep learning techniques, specifically neural networks such as recurrent neural networks (RNNs), attention mechanisms, and transformer architectures, to treat translation as a sequence-to-sequence learning problem, resulting in more fluent, contextually aware translations. Recent advancements encompass the application of large language models (LLMs) and

generative AI, which can be optimized for translation tasks and demonstrate potential in addressing cultural subtleties and creative language, especially when integrated with distinctive prompting strategies and adapter-based tuning. Moreover, hybrid and multi-strategy systems employ semantic analysis, context modeling, and fuzzy logic to enhance translation precision. The quality of AI translation is often assessed using metrics such as BLEU and METEOR, though contemporary methods employ AI to evaluate translation accuracy and stylistic integrity. Notwithstanding these advancements, challenges remain in translating under-resourced languages and expressing cultural or emotional subtleties; nonetheless, continuous research is expanding the capabilities of AI translation (Kodali, Upreti & Boppana, 2024 & Elkins, 2024).

This research aims to use AI-based translation approaches to teach students how to translate effectively. On one hand, neural machine translation (NMT) and other digital platforms are examples of AI-powered translation systems that have been shown in numerous studies to be helpful teaching tools for language acquisition and translation education. Particularly for advanced learners who can evaluate AI-generated outputs critically, NMT tools have been demonstrated to develop both productive (speaking, writing) and receptive (reading, listening) language abilities as well as translation-related mediation skills. Compared with conventional approaches, AI-based translation technologies dramatically improve readability and translation accuracy, while also boosting teacher satisfaction and student performance. Although a balanced approach with traditional methods is advised to avoid over-reliance, AI tools also improve translation accuracy and have a favourable impact on learners' motivation and emotional engagement (Klimova, Pikhart, Benites, Lehr & Sanchez-Stockhammer, 2022; Yuxiu, 2024; Kruk & Kałużna, 2024).

On the other hand, the majority of teachers are cautiously hopeful about the use of AI and machine translation in the classroom and acknowledge translation as a crucial pedagogical tool. Teachers perceive value in using AI tools for independent study, especially for advanced learners and for supporting children with additional linguistic challenges, even though there are still issues

around academic integrity and passive learning. The need for guided pedagogical integration is highlighted by the fact that students commonly use web-based machine translation tools to support their studies. However, they may find it challenging to evaluate AI-generated translations critically and have little faith in the veracity of the result (Tasmedir, Lopez, Sata & Riches, 2023 & Khasawneh & Shawaqfeh, 2024).

Moreover, Real-time collaboration, quicker creation of instructional materials, and automated grading, which cuts down on assessment time and gives students instant feedback, are all made possible by AI and digital translation technologies. Additionally, these resources provide deeper contextual understanding, real-time feedback, and customised learning opportunities, all of which support better language learning and translation. Although ease of use remains a hurdle, senior lecturers and older educators find AI tools valuable for effective teaching, underscoring the need for user-friendly platforms and targeted training. Notwithstanding the advantages, there are also issues, such as institutional and technological constraints, concerns about

becoming overly dependent on technology, and the need for critical engagement with AI outputs. Both educators and learners emphasize the importance of understanding the advantages and disadvantages of AI translation systems and support their careful incorporation into instructional settings (Karasaliu, 2024).

As a result, Artificial Intelligence techniques are revolutionising current language teaching by improving language competency, personalising learning, and expediting assessment and training. AI excels at improving abilities such as pronunciation, grammar, and vocabulary while also enabling adaptable, efficient, and enjoyable learning experiences. However, it cannot entirely replace the human parts of teaching, such as developing cultural sensitivity and critical thinking. AI techniques greatly improve key language skills—listening, speaking, reading, and writing—through tailored feedback, particularly in pronunciation, grammar, and vocabulary development. Personalised Learning: Adaptive learning systems and AI chatbots tailor information and practice to individual student needs, accommodating diverse learning styles and enhancing engagement. AI Automated

Assessment: AI allows for rapid, objective, and scalable assessment of language skills, providing fast feedback and facilitating both formative and summative evaluations. Teaching Support: Artificial intelligence (AI) helps teachers prepare materials, grade assignments, guide student practice, and facilitate classroom interactions, thereby increasing teaching efficiency (Sun, Anbarasan, Kumar & Kumar, 2020 & Qassrawi, Karasneh, 2025).

However, AI as Human Skills has limitations and challenges: AI struggles to learn cultural sensitivity, emotional intelligence, and critical thinking, which are essential for communication and relationships. Implementation Obstacles: Teachers struggle with technical support, training, and curriculum alignment. Data privacy, permission, and bias are AI ethics challenges (Kovalenko & Baranivska, 2024; Zhou & Hou, 2025). To conclude, AI technologies make language learning more engaging, efficient, and personalised. They should supplement human educators to develop critical and interpersonal skills. Careful implementation, teacher training, and ethical and pedagogical attention are needed for successful integration.

In addition, research regularly reveals that AI translation systems, such as DeepL and other AI-powered platforms, can improve EFL (English as a Foreign Language) learners' writing skills and vocabulary acquisition. Studies show that utilising AI translation and writing tools improves grammar, sentence structure, vocabulary enrichment, and general writing proficiency, with students benefiting from context-sensitive translations and instant feedback on their work (Payung & Sukarno, 2025 & Rafida, Suwandi & Ananda,2024). AI-powered applications such as Duolingo and integrated platforms have been shown to improve vocabulary learning efficiency and outcomes regardless of gender, and they are suggested for inclusion in the EFL curriculum. Meta-analyses and experimental studies demonstrate that AI chatbots and writing tools improve the structure, coherence, and content quality while also enhancing motivation and engagement. However, significant concerns persist regarding over-reliance on AI, diminished creativity, and difficulties with authenticity and plagiarism, underscoring the importance of a balanced integration with effective

teaching practices. Overall, the evidence demonstrates that, when strategically integrated, AI translation and writing tools provide significant benefits to EFL learners' writing progress and vocabulary increase, while also requiring ongoing study and careful pedagogical application (Song & Song, 2023 & Wang, Wu, Chen, Wang, Li & Wang, 2024).

The research questions

1. To what extent do AI-based translation tools (specifically DeepL, ChatGPT, and Google Gemini) foster (EFL) writing skills among students at Al Madinah International University?

2. To what extent do AI-based translation tools (specifically DeepL, ChatGPT, and Google Gemini) foster (EFL) vocabulary skills among students at Al Madinah International University?

The experts agreed on the validity, suitability, and soundness of the test items.

b) Internal consistency validity method

Methods

The test was applied to a pilot sample of Al Madinah international students, numbering 15 male and female students, to determine the following:

a) Calculating the Validity of the Test:

The subsequent methodology employed to ascertain the test's validity was the Method of Expert Validation.

- The exam's validity was assessed by expert validation, which included presenting it to a panel of experts for their evaluation of the adequacy of instructions given to students for accurately answering the test.

- The scientific and linguistic validity of the items.

- Appropriateness of the goods for the selected students.

- The appropriateness of each inquiry for assessing the intended skill.

- Each question fulfills its intended purpose.

The researcher computed the Pearson correlation coefficient between students' scores on each skill and the test's overall score using SPSS V.18 to determine the test's

internal consistency validity. The validity coefficients for the writing skills test's subskills are displayed in the following table:

Table 1 shows the correlation coefficient between each skill score and the total score of the writing skills test (N = 20).

Skills	Content & Ideas	Organization & Cohesion	Vocabulary & Word Choice	Grammar & Sentence Structure	Mechanics & Spelling
Correlation	0.743**	0.637**			0.725**

**. Correlation is significant at the 0.01 level

The previous table shows that all of the correlation coefficients are significant at the 0.01 level. This means that the test's sub-skills are internally consistent.

c) Discriminant Validity Method

Discriminant validity was assessed to confirm the writing skills test's capacity to discriminate. The pilot sample's lower percentiles accounted for 27% of the low scores, while the upper percentiles (20 students) accounted for 27% of the high scores. The significance of the differences between these two groups was evaluated using the non-parametric Mann-Whitney U test. The findings of the z-value and mean rank discrepancies between the two groups are displayed in the following table. The following are the outcomes:

Table (2)

The differences between the mean ranks and the z-value for the writing skills test between the two groups.

Group	N	Mean Rank	Sum of Ranks	Z-Value	Sig
High-Level Group	6	9.50	57.00	2.903	0.01
Low-Level Group	6	3.50	21.00		

The chart clearly demonstrates a statistically significant difference at the 0.01 level between the two groups, indicating good discriminative validity of the test.

d) Test   Reliability   Calculation

The reliability of the test was calculated using the following methods:

1. **Cronbach's** **Alpha** **Method:**

   For the writing skills test, the Cronbach's alpha coefficient was computed and was 0.797. Given the high value, it may be concluded that the test is highly reliable.

2. **Test-Retest** **Reliability** **Method:**

   After the pilot sample kids had taken the test, the same sample was given the test again two weeks later. Using SPSS (V. 18), the Pearson correlation coefficient between the student scores on the two administrations was computed. The following table displays the correlation coefficient between the two administrations for each skill and the test as a whole:

Table *(3)*

Reliability Coefficients for the Writing Skills Test (N = 30)

Skills	Content & Ideas	Organization & Cohesion	Vocabulary & Word Choice	Grammar & Sentence Structure	Mechanics & Spelling	Overall Test
Correlation	0.792**	0.726**	0.810**	0.751	0.803*	0.907**

It is clear from the above table that all correlation coefficients are high and are significant at the 0.01 level, indicating the reliability of each primary skill and the test as a whole.

e) Test Duration Calculation:

The researcher determined that 45 minutes is a reasonable amount of time for the test by averaging the times taken by each sample member.

2. Pilot Testing of the Vocabulary Acquisition Skills Test:

The test was administered to a pilot sample of 20 students at Al Madinah International University. The purpose was to determine the following:

a) Calculating the Validity of the Test:

The validity of the test was assessed using the following methods:

Judges' Validity Method

A panel of experts was presented with the test to assess its validity. They were asked to comment on the following: o The items' scientific and linguistic validity. The sufficiency of the instructions given to students for appropriately answering .The appropriateness of the items for the sample

students, each question's fitness for the skill it seeks to assess, and whether each question accomplishes its intended goal. The validity, appropriateness, and integrity of the test and its items were all agreed upon by the experts.

Internal Consistency Validity Method

a) The researcher measured the internal consistency between each subskill score and the overall score of the primary skill to which it belongs using SPSS V.18 to determine internal consistency validity. b) The degree of internal consistency between the total test score and the scores for each paramount ability.

(1) Internal consistency between each sub-skill score and the overall score of the main skill

Pearson correlation coefficients between students' scores on each subskill and the overall score of the primary skill to which that subskill belongs were used to determine the internal consistency. The validity coefficients of the EFL vocabulary test's subskills are displayed in the following table:

Table (4)   The correlation coefficient between the score of each sub-skill and the total score of the main skill to which that sub-skill belongs for the Vocabulary Acquisition Skills Test (N = 20).

No	Main Skills	Sub-Skills	Correlation
1	Understanding Vocabulary in Context	Inferring meaning from the surrounding text (words, sentences, paragraphs).	0.642**
2		Recognising the general gist of an unfamiliar word based on topic and tone.	0.723**
3		Identifying the part of speech of a word based on its position in a sentence.	0.542*
4		Understanding abstract vocabulary and nuanced meanings related to complex topics (e.g., societal issues, ethics, scientific concepts).	0.642**

5	Recognising and Interpreting Common Word Patterns	Identifying and understanding common prefixes (e.g., *un, re, dis, pre, inter*) and suffixes (e.g., *-tion, -ment, -able, -ful, -less, -ize, ly*) to deduce word meaning.	0.827**
6		Recognising the base form of a word from its derivations (e.g., *negotiation → negotiate*).	0.487*
7		Understanding common collocations (words that frequently go together) when encountered in reading or listening (e.g., *heavy rain*, *make a decision*, *take responsibility*).	0.577**
8		Interpreting common phrasal verbs, including those with multiple meanings (e.g., *look up*, *put off*, *take on*).	0.755**
9		Understanding the meaning of common idioms and figurative expressions (e.g., *beat around the bush*, *once in a blue moon*).	0.537*
10	Discriminating Vocabulary Nuances	Differentiating between synonyms with slightly different meanings or connotations (e.g., *old* vs. *ancient* vs. *elderly*).	0.449*
11		Recognising words with multiple meanings based on context (polysemy).	0.843**
12		Identifying the register (formal, informal, neutral) of vocabulary used in a text or speech.	0.674**
13	Using Vocabulary Appropriately in Speaking	Employing a wide range of vocabulary to express detailed and nuanced ideas on a variety of B2 topics (e.g., presenting arguments, discussing abstract concepts, expressing opinions).	0.782**
14		Accurately using common collocations in spontaneous speech.	0.780**

15		Integrating frequently used phrasal verbs and idioms naturally into conversations.	0.736**
16		Selecting appropriate vocabulary for different registers and social situations.	0.660**
17		Actively trying to incorporate newly learned words into daily conversations.	0.883**
18		Applying a broad vocabulary to write clear, coherent, and well-structured texts on complex subjects.	0.745**
19		Accurately using word families and derivations (e.g., correctly forming nouns, verbs, adjectives, adverbs from a root word).	0.829**
20	Using Vocabulary Accurately in Writing	Consistently use correct collocations in written assignments.	0.553*
21		Incorporating appropriate phrasal verbs and idioms to enrich written expression.	0.714**
22		Maintaining a consistent and appropriate register throughout a written piece (e.g., formal for essays, informal for personal emails).	0.708**
23		Using a variety of synonyms to avoid repetition and enhance lexical richness.	0.788**

*. Correlation is significant at the 0.05 level & **. Correlation is significant at the 0.01 level.

It is evident from the previous table that all correlation coefficients are significant at the 0.05 and 0.01 levels, indicating internal consistency validity among the test's sub-skills.

(2) Calculating the internal consistency between each main skill score and the overall test score

The Pearson correlation coefficient between each major skill score and the overall test score was used to determine the internal consistency. The primary skills' validity coefficients for the Vocabulary Acquisition Skills Test are displayed in the following table:

Table (5)

The validity coefficients of the main skills for the Vocabulary Acquisition Skills Test

Skills	Understanding Vocabulary in Context	Recognising and Interpreting Common Word Patterns	Discriminating Vocabulary Nuances	Using Vocabulary Appropriately in Speaking	Using Vocabulary Accurately in Writing
Correlation	0.720**	0.571*	0.542*	0.598**	0.556*

*Note: Correlation is significant at the 0.05 level; ** at the 0.01 level.

It is clear from the previous table that all correlation coefficients are significant at the 0.05 and 0.01 levels, indicating the validity of the test's main skills.

Discriminant Validity Method

Discriminant validity was assessed to confirm the Vocabulary Acquisition Skills Test's discriminant validity. From the 20 students in the pilot sample, the top 27% of scores were selected from the high end, and the bottom 27% from the low end. The significance of the differences between these two groups was evaluated using the Mann-Whitney U test. The following table shows the results of the differences between the mean ranks and the z-value between the two groups.

Table (6)

The results of the differences between the mean ranks and the z-value between the two groups

Group	N	Mean Rank	Sum of Ranks	Z-Value	Sig
High-Level Group	6	9.50	57.00	2.929	0.01
Low-Level Group	6	3.50	21.00		

The table shows a statistically significant difference at the 0.01 level between the two groups, indicating high discriminant validity.

3. **Test   Reliability   Calculation**

The reliability of the test was calculated using the following method:

Cronbach's   Alpha   Method:

Cronbach's alpha coefficients were computed for both the entire test and its primary skills. The table below displays the findings.

Table(7)

Cronbach's alpha coefficient for the vocabulary acquisition skills test (N = 20)

skills	Understanding vocabulary in context	recognising and interpreting common word patterns	discriminating vocabulary nuances	using vocabulary appropriately in speaking	using vocabulary accurately in writing	overall test
Cronbach's alpha	0.711	0.732	0.702	0.817	0.783	0.836

The preceding table shows that the measure has a high degree of reliability, as indicated by Cronbach's alphas.

4. **Test-Retest Reliability Method:**

The test was given to the pilot sample students and subsequently re-administered after a two-week gap. The Pearson correlation coefficient was computed between the scores of students from both administrations, utilising SPSS (V. 18). The correlation coefficients for each skill and the overall test between the two administrations are presented in the subsequent table:

Table(8)

Reliability coefficient for the vocabulary acquisition skills test (N = 20).

skills	Understanding vocabulary in context	recognising and interpreting	discriminating vocabulary nuances	using vocabulary	using vocabulary	overall test

		common word patterns		appropriately in speaking	accurately in writing	
correlation	0.723**	0.762**	0.803**	0.722**	0.794**	0.931**

The preceding table makes it evident that every correlation coefficient is high and significant at the 0.01 level, demonstrating the dependability of both the test overall and each key skill individually.

Calculating Test Duration:

The test duration was calculated by averaging the times of all sample members. The best duration for the exam, according to the researcher, is one hour.

Selecting the Study Sample and Dividing it into Two Equivalent Groups

Two student groups were used in the study: one was experimental and consisted of 15 students who were studied using the AI translation-based approach; the other was control and consisted of 15 students who were studied using the conventional teaching approach. As displayed in the following table:

Table (9)

The number of the pre- and post-groups

Group	Control	experimental	Total
Pre	15	15	**30**
Post	15	15	**30**

The Experimental Design of the Study:

This study examines the effect of one or more experimental factors on one or more dependent factors and falls under the category of quasi-experimental investigations. As a result, a particular experimental design was employed: a pretest-posttest design with two groups, a control group and an experimental group.

4. Procedures of the Study

1- Equivalence of the Study Groups

The most important extraneous variable that could affect the dependent variables must be controlled to examine the impact of AI translation tools on EFL writing and vocabulary acquisition skills. This guarantees that the independent variable is the only cause of any changes in these variables observed. These factors are:

a) Social, Cultural, and Economic Level

The two study groups come from the same social setting. This suggests that their social, cultural, and economic spheres are comparable. Consequently, in terms of this variable, the two groups can be regarded as equal.

b) The EFL writing proficiency of the pupils

The mean rank scores of the experimental and control groups in the pre-test of the full EFL Writing test, as well as for each primary ability, were compared using the Mann-Whitney U test to ensure the groups are equivalent in EFL Writing. This is shown in the table below:

Table *(10)*

Mann-Whitney Test between the mean rank scores of the experimental and control groups. (N1 = N2 15)

Main Skills	Group	Mean Rank	Sum of Ranks	U-value	Z-value	α Sig
Content & Ideas	Experimental	16.90	253.50	91.5	1.027	**0.304**
	Control	14.10	211.50			
Organization & Cohesion	Experimental	15.50	232.50	112.5	0.000	**1.00**
	Control	15.50	232.50			
Vocabulary & Word Choice	Experimental	17.00	255.00	90.0	1.077	**0.281**
	Control	14.00	210.00			
Grammar & Sentence Structure	Experimental	15.50	232.50	112.5	0.000	**1.000**
	Control	15.50	232.50			

Mechanics & Spelling	Experimental	17.00	255.00	90.0	1.087	**0.277**
	Control	14.00	210.00			
All Over the Test	**Experimental**	**17.63**	**264.50**	**80.5**	**1.371**	**0.170**

The aforementioned table indicates that, for both each primary skill and the test overall, the value of "Z" is not statistically significant at the level $\alpha < 0.05$. This shows that before the study experiment's deployment, the two groups' EFL writing proficiency was equal in both overall and in each of the major skills.

a) Student Levels in Vocabulary Acquisition Skills

The Mann-Whitney U test was used to compare the mean rank scores of the experimental and control groups in the pre-test of Vocabulary Acquisition Skills as a whole and for each key skill. This was done to make sure that the two study groups were the same in this area. The table below shows this:

Table **(11)**

Mann-Whitney Test between the mean rank scores of the experimental and control groups(N1 = N2 = 15)

Main Skills	Group	Mean Rank	Sum of Ranks	U-value	Z-value	α Sig
Understanding Vocabulary in Context	Experimental	14.50	217.5	97.50	0.720	**0.472**
	Control	16.50	247.5			
Recognising and Interpreting Common Word Patterns	Experimental	14.20	213.0	93.00	0.960	**0.337**
	Control	16.80	252.0			
Discriminating Vocabulary Nuances	Experimental	14.00	210.0	90.00	1.077	**0.281**
	Control	17.00	255.0			
	Experimental	17.50	262.5	82.50	1.466	**0.143**

Using Vocabulary Appropriately in Speaking	Control	13.50	202.5			
Using Vocabulary Accurately in Writing	Experimental	13.43	201.5	81.50	1.503	**0.133**
	Control	17.57	263.5			
All Over the Test	**Experimental**	**13.93**	**209.0**	**89.00**	**1.020**	**0.308**
	Control	**17.07**	**256.0**			

According to the preceding table, the value of "Z" for each major skill and for the test overall is not statistically significant at the $\alpha < 0.05$ level. This suggests that before the study experiment's deployment, the two groups were equal in terms of their overall and specific vocabulary acquisition skills.

Discussion and interpretation of the results

1- Presentation and discussion of results related to the first hypothesis:

The study's first hypothesis says, "There is a statistically significant difference between the mean ranks of the scores of the experimental and control groups in the post-test of the overall writing skills and its main components in favour of the experimental group students." The Mann-Whitney Test was utilised to assess the validity of this hypothesis by calculating the difference in the mean ranks of the post-test scores from the experimental and control groups for overall writing skills and its primary components. As shown in the table below, we also found that the experimental treatment improved writing skills.

Table 12

The effect size of the experimental treatment on writing skills

Main Skills	Group	Mean Rank	Sum of Ranks	U-value	Z-value	α Sig	) r_{rb} (
Content & Ideas	Experimental	22.33	335.00	10.00	4.440	0.000	**0.911**
	Control	8.67	130.00				
Organization & Cohesion	Experimental	22.77	341.50	3.50	4.684	0.000	**0.969**
	Control	8.23	123.50				

Vocabulary & Word Choice	Experimental	23.00	345.00	0.00	4.853	0.000	**1**
	Control	8.00	120.00				
Grammar & Sentence Structure	Experimental	23.00	345.00	0.00	4.827	0.000	**1**
	Control	8.00	120.00				
Mechanics & Spelling	Experimental	23.00	345.00	0.00	4.853	0.000	**1**
	Control	8.00	120.00				
All Over the Test	**Experimental**	**23.00**	**345.00**	**0.00**	**4.709**	**0.000**	**1**
	Control	**8.00**	**120.00**				

The preceding table makes this evident: - The mean ranks of the experimental and control groups' scores in the post-application of the writing skills test as a whole, and its main skills differ statistically significantly at the level ($\alpha \leq 0.01$) in favour of the experimental group's scores. This suggests that the study's initial hypothesis has been accepted. For both overall and individual primary writing skills, the experimental treatment's effect size ranged from 0.911 to 1.00, which is high and exceeds 0.9, indicating a powerful effect.

2- Presentation and discussion of the results related to the second hypothesis:

To evaluate the validity of the second hypothesis of the study, which asserts that "there is a statistically significant difference between the mean ranks of the experimental group's scores in the pre- and post-applications of the writing skills test as a whole and its main skills in favour of the post-application," the significance of the differences between the mean ranks of the experimental group's scores in the pre- and post-applications of the writing skills test as a whole and its main skills was assessed using the Wilcoxon Signed Ranks Test. The table below shows the effect size of the experimental treatment on writing skills, which was also calculated.

Table (13)

The effect size of the experimental treatment on the writing skills

Main Skills	Ranks	N	Mean Rank	Sum of Ranks	Z-value	α Sig	(r_{rb})
Content & Ideas	Negative Ranks	0	0.00	0.00	3.638	0.000	**0.750**

	Positive Ranks	14	7.50	105.00			
	Ties	1					
Organization & Cohesion	Negative Ranks	0	0.00	0.00	3.690	0.000	1
	Positive Ranks	15	8.00	120.00			
	Ties	0					
Vocabulary & Word Choice	Negative Ranks	0	0.00	0.00	3.771	0.000	1
	Positive Ranks	15	8.00	120.00			
	Ties	0					
Grammar & Sentence Structure	Negative Ranks	0	0.00	0.00	3.771	0.000	1
	Positive Ranks	15	8.00	120.00			
	Ties	0					
Mechanics & Spelling	Negative Ranks	0	0.00	0.00	3.690	0.000	1
	Positive Ranks	15	8.00	120.00			
	Ties	0					
All Over the Test	**Negative Ranks**	**0**	**0.00**	**0.00**	**3.623**	**0.000**	**1**
	Positive Ranks	**15**	**8.00**	**120.00**			
	Ties	**0**					

It is evident from the above table that the mean rankings of the experimental group students' results on the writing skills test as a whole and its primary skills differ statistically significantly at the level ($\alpha < 0.01$), favouring the post-application. This suggests that the study's second hypothesis has been accepted.

- Except for the skill Content & Ideas, where the effect size was 0.750, suggesting a strong effect of the experimental treatment in this skill, the experimental treatment had a substantial effect on the writing skills as a whole and in each of its main skills, reaching (1).

3- Presentation and discussion of the results for the third hypothesis:

To test the validity of the third hypothesis of the study, which states that "There is a statistically significant difference between the mean ranks of the scores of the experimental and

control groups in the post-application test of vocabulary acquisition skills as a whole and its main skills in favor of the experimental group students," the Mann-Whitney Test was calculated to determine the difference between the mean ranks of the scores of the experimental and control groups in the post-application test of vocabulary acquisition skills as a whole and its main skills. The effect size of the experimental treatment on vocabulary acquisition was also calculated, as shown in the following table.

Table (14)

The effect size of the experimental treatment on vocabulary acquisition skills

Main Skills	Group	Mean Rank	Sum of Ranks	U-value	Z-value	α Sig	(r$_{rb}$)
Understanding Vocabulary in Context	Experimental	23.00	345.00	0.000	4.871	0.000	**1**
	Control	8.00	120.00				
Recognising and Interpreting Common Word Patterns	Experimental	22.70	340.50	4.500	4.664	0.000	**0.960**
	Control	8.30	124.50				
Discriminating Vocabulary Nuances	Experimental	20.40	306.00	39.00	3.358	0.000	**0.653**
	Control	10.60	159.00				
Using Vocabulary Appropriately in Speaking	Experimental	22.40	336.00	9.00	4.462	0.000	**0.920**
	Control	8.60	129.00				
Using Vocabulary Accurately in Writing	Experimental	22.80	342.00	3.00	4.646	0.000	**0.973**
	Control	8.20	123.00				
All Over the Test	**Experimental**	**23.00**	**345.00**	**0.000**	**4.706**	**0.000**	**1**
	Control	**8.00**	**120.00**				

It is clear from the previous table:

- There is a statistically significant difference at the level ($\alpha \leq 0.01$) between the mean ranks of the vocabulary acquisition skills test scores for the experimental and control groups in the

post-application, in favour of the experimental group students' scores. This indicates the acceptance of the third hypothesis of the research.

- The effect size of the experimental treatment on the vocabulary acquisition skills as a whole and on each of its main skills ranged between (0.920 – 1.00), which is a significant value and greater than (0.9), indicating a powerful effect in the test as a whole and in all its skills, except for the skill Discriminating Vocabulary Nuances, where the effect size was (0.653), indicating a moderate effect in this skill.

4- **Presentation and discussion of the results related to the fourth hypothesis**

To assess the validity of the study's fourth hypothesis, which reads, "There is a statistically significant difference between the mean ranks of the experimental group students' scores in the pre- and post-applications of the vocabulary acquisition skills test as a whole and its main skills in favour of the post-application," the Wilcoxon Signed Ranks Test was used to calculate the significance of the differences between the mean ranks of the experimental group students' scores in the pre- and post-applications of the vocabulary acquisition skills test as a whole and its main skills. The following table shows how the experimental treatment's effect size on language learning abilities was determined.

Table (15)

The experimental treatment effect size on vocabulary acquisition skills

Main Skills	Ranks	N	Mean Rank	Sum of Ranks	Z-value	α Sig	(r_{rb})
Understanding Vocabulary in Context	Negative Ranks	0	0.00	0.00	3.638	0.000	**0.750**
	Positive Ranks	15	8.00	120.00			
	Ties	0					
Recognising and Interpreting Common Word Patterns	Negative Ranks	0	0.00	0.00	3.690	0.000	**1**
	Positive Ranks	15	8.00	120.00			
	Ties	0					
	Negative Ranks	0	0.00	0.00	3.771	0.000	**0.30**

Discriminating Vocabulary Nuances	Positive Ranks	12	6.50	78.00			
	Ties	3					
Using Vocabulary Appropriately in Speaking	Negative Ranks	0	0.00	0.00	3.771	0.000	**1**
	Positive Ranks	15	8.00	120.00			
	Ties	0					
Using Vocabulary Accurately in Writing	Negative Ranks	0	0.00	0.00	3.690	0.000	**1**
	Positive Ranks	15	8.00	120.00			
	Ties	0					
All Over the Test	**Negative Ranks**	**0**	**0.00**	**0.00**	**3.623**	**0.000**	**1**
	Positive Ranks	**15**	**8.00**	**120.00**			
	Ties	**0**					

It is clear from the above table:

- There is a statistically significant difference at the level ($\alpha \leq 0.01$) between the average ranks of scores of the experimental group students in both the pre- and post-application of the vocabulary acquisition skills test as a whole and its main skills in favour of the post-application. This indicates the acceptance of the fourth hypothesis of the research.

- The effect size of the experimental treatment on vocabulary acquisition skills as a whole and in each of its main skills reached (1), indicating a considerable effect size of the experimental treatment on writing skills as a whole and its sub-skills, except for the skill of Discriminating Vocabulary Nuances, where the effect size was (0.300), indicating a weak effect of the experimental treatment in this skill.

The discussion

The outcomes of this study, indicating statistically significant improvement in EFL writing and vocabulary acquisition among students utilising AI-

based translation software (DeepL, ChatGPT, and Google Gemini), offer significant revelations into the evolving role of technology in language learning. The results not only highlight the practical benefits of such tools but also resonate with, and occasionally rephrase, existing academic research on translation in foreign language learning.

The debate surrounding the use of translation in language acquisition, foregrounded by Yulita (2021), has long existed: whether it can enhance learning or suffocate it through excessive dependence on the mother tongue. The clear-cut assertion of enhanced writing and vocabulary skills resoundingly supports the stance that, if utilised effectively, translation can be a potent pedagogic tool. While Yulita's review documented mixed findings regarding the impact of proficiency level on the benefits of translation, the overall positive impacts of the present university setting suggest the broad generalizability of AI-enabled translation aids across various student capabilities, providing further empirical support for this controversy.

Moreover, the positive impact of vocabulary acquisition identified in the present research strongly supports Tukan's (2024) arguments. Tukan's study emphasised that translation activities, particularly in promoting linguistic tools, lead to significant improvements in vocabulary, retention, and writing accuracy and fluency. The study develops this consciousness by showing that advanced AI tools like DeepL, ChatGPT, and Google Gemini not only facilitate such gains but also reinforce greater student confidence and autonomy in word acquisition. This convergence of perceived use and actual benefit is the primary link Tukan identified between students' confidence in translation and its practical use in pedagogical settings. To what extent AI makes support available and sophisticated likely assists in building this confidence and enables independent learning.

Thus, the findings of the current study directly address and harmonise with contemporary reevaluation of the pedagogical function of translation. While Kalyani (2023) justly criticises the defects of the traditional Grammar-Translation Method (GTM), which far too often restricts language use to the ordinary and encourages rote memorisation, the research supports the more equitable viewpoint articulated by Damacena

(2024) and Naghiyeva (2025). These writers confirm that translation, when "used wisely," can be a valuable source for analytical thinking, intercultural competence, and learning complex linguistic structures. The evidence that AI-based translation tools can be "useful additional resources" reflects their appeal as an "added resource" that complements contextual and communicative learning rather than a standard, individual technique. By allowing for dynamic understanding and production, AI translation software goes beyond the limitations of the GTM, demonstrating how technology can ensure a "wise application" of translation that supports rather than impedes the development of natural language ability and communicative competence in EFL learners.

Conclusion

This study clearly shows that AI-based translation tools improve both EFL writing and vocabulary acquisition skills among university students. The findings demonstrate that incorporating technologies such as DeepL, ChatGPT, and Google Gemini into language learning not only yields quantitative improvements in these essential competencies but also boosts students' confidence and autonomy in their learning journey. While the impact was significant across most skills, specific areas, such as "Content & Ideas" in writing and "Discriminating Vocabulary Nuances" in vocabulary, also showed favourable, albeit variable, benefits. These findings strongly support the strategic use of AI-based translation tools as valuable supplementary resources in EFL teaching, with significant implications for future curriculum development and technological improvements in language instruction.

References:

Albelihi, H. (2022). Lexical and cohesive links in EFL learners' writing: Exploring the use of task-based language teaching, 7. https://doi.org/10.3389/feduc.2022.996171.

Balle, S., & Olsen, M. (2024). From words to text – academic writing in EFL classrooms. Nordic Journal of Language Teaching and Learning. https://doi.org/10.46364/njltl.v11i2.1149.

Chomicz, M. (2025). Enhancing EFL writing skills for adult Deaf and hard-of-hearing individuals. Frontiers in Education. https://doi.org/10.3389/feduc.2024.1504503.

Damacena, B. (2024). THE ROLE OF TRANSLATION IN FOREIGN LANGUAGE

TEACHING. International Journal of Language, Linguistics, Literature, and Culture. https://doi.org/10.59009/ijlllc.2024.0056.

Dhuli, R., Lamo, P., & Larsari, V. (2023). An Analysis of the Significance of Vocabulary in Fostering ESL/EFL Students' Writing Skills: An Empirical Study. International Journal of Contemporary Studies in Education (IJ-CSE). https://doi.org/10.56855/ijcse.v2i1.252.

Elkins, K. (2024). In search of a translator: using AI to evaluate what's lost in translation. Frontiers in Computer Science. https://doi.org/10.3389/fcomp.2024.144402.

Jin, Z., & Webb, S. (2021). Does writing words in notes contribute to vocabulary learning? Language Teaching Research. https://doi.org/10.1177/13621688211062184

Jurado, A. (2023). The Impact of Creative Writing Activity on EFL Learners' Non-Core Vocabulary Acquisition. Journal of English Language Teaching and Linguistics. https://doi.org/10.21462/jeltl.v8i3.1189.

Kakihana, S. (2024). Does the word-writing strategy in vocabulary learning increase learning outcomes or mislead learners into overestimation? International Journal of Applied Linguistics. https://doi.org/10.1111/ijal.12544.

Kalyani, L. (2023). Translation as a Learning Method in English Language Teaching. International Journal

of Multidisciplinary Research in Arts, Science and Technology. https://doi.org/10.61778/ijmrast.v1i3.12.

Karasaliu, A. (2024). Comparative Analysis of Digital Translation Tools. The Eurasia Proceedings of Science Technology Engineering and Mathematics. https://doi.org/10.55549/epstem.1593218.

Kartika-Ningsih, H. (2023). Implementing a reading-to-learn programme in EFL bilingual teaching. ELT Journal. https://doi.org/10.1093/elt/ccad042.

Khasawneh, M., & Shawaqfeh, A. (2024). Breaking Traditional Boundaries in Translation Pedagogy; Evaluating How Senior Lecturers Have Incorporated Digital Tools to Enhance Translation Teaching. World Journal of English Language. https://doi.org/10.5430/wjel.v14n4p154.

Klimova, B., Pikhart, M., Benites, A., Lehr, C., & Sanchez-Stockhammer, C. (2022). Neural machine translation in foreign language teaching and learning: a systematic review. Education and Information Technologies, 28, 663-682. https://doi.org/10.1007/s10639-022-11194-2.

Kodali, R., Upreti, Y., & Boppana, L. (2024). A Novel Approach to Generative AI Translation. TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON), 698-702. https://doi.org/10.1109/TENCON61640.2024.10902899.

Kovalenko, I., & Baranivska, N. (2024). Integrating Artificial Intelligence in English Language Teaching: Exploring the Potential and Challenges of Ai Tools in

Enhancing Language Learning Outcomes and Personalized Education. European Socio-Legal & Humanitarian Studies. https://doi.org/10.61345/2734-8873.2024.1.9.

Kruk, M., & Kałużna, A. (2024). Investigating the Role of AI Tools in Enhancing Translation Skills, Emotional Experiences, and Motivation in L2 Learning. European Journal of Education. https://doi.org/10.1111/ejed.12859.

Lee, Y. (2021). Still taboo? Using machine translation for low-level EFL writers. ELT Journal. https://doi.org/10.1093/elt/ccab018.

Martín, C., Hirsu, L., Gonzales, L., & Alvarez, S. (2019). Pedagogies of Digital Composing through a Translingual Approach. Computers and Composition. https://doi.org/10.1016/J.COMPCOM.2019.02.007.

Mondal, S., Zhang, H., Kabir, H., Ni, K., & Dai, H. (2023). Machine translation and its evaluation: a study. Artificial Intelligence Review, 56, 10137-10226. https://doi.org/10.1007/s10462-023-10423-5.

Naghiyeva, G. (2025). Revamping Traditional Methods: Evaluating the Grammar-Translation Method in Modern Language Teaching. Acta Globalis Humanitatis et Linguarum. https://doi.org/10.69760/aghel.02500111.

Payung, N., & Sukarno, S. (2025). Enhancing EFL Students' Writing Skills: The Impact of Deepl Translation. Journal of Languages and Language Teaching. https://doi.org/10.33394/jollt.v13i1.12301.

Qassrawi, R., & Karasneh, S. (2025). Redefinition of human-centric skills in language education in the AI-driven era. Studies in English Language and Education. https://doi.org/10.24815/siele.v12i1.43082.

Rafida, T., Suwandi, S., & Ananda, R. (2024). EFL Students' Perception in Indonesia and Taiwan on Using Artificial Intelligence to Enhance Writing Skills. Jurnal Ilmiah Peuradeun. https://doi.org/10.26811/peuradeun.v12i3.1520.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1260843.

Sun, Z., Anbarasan, M., Kumar, D., & Kumar, P. (2020). Design of online intelligent English teaching platform based on artificial intelligence techniques. Computational Intelligence, 37, 1166 - 1180. https://doi.org/10.1111/coin.12351.

Tasmedir, S., Lopez, E., Satar, M., & Riches, N. (2023). Teachers' perceptions of machine translation as a pedagogical tool. The JALT CALL Journal. https://doi.org/10.29140/jaltcall.v19n1.24.

Tu, M., Q., & Jiang, L. (2024). Exploring EFL vocabulary learning through the story continuation writing task. Australian Review of Applied Linguistics. https://doi.org/10.1075/aral.24036.tu.

Tukan, F. (2024). The Difficulties and Strategies of Efl Students in Improving Their English Skills. IJIET (International Journal of Indonesian Education and Teaching). https://doi.org/10.24071/ijiet.v8i1.3375.

Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J., & Wang, L. (2024). Empirical Assessment of AI-Powered Tools for Vocabulary Acquisition in EFL Instruction. IEEE Access, 12, 131892-131905. https://doi.org/10.1109/ACCESS.2024.3446657.

Wei, B. (2024). Advancements and challenges in AI-driven creative translation: A comprehensive analysis. Applied and Computational Engineering. https://doi.org/10.54254/2755-2721/82/20240937.

Wu, L., & Halim, H. B. A. (2024, March). Task complexity and foreign language writing emotions as predictors of EFL writing performance. In Frontiers in Education (Vol. 9, p. 1323843). Frontiers Media SA.

Yulita, D. (2021). The correlation of English proficiency level and translation strategies used by Indonesian EFL learners. LLT Journal: A Journal on Language and Language Teaching, 24(1), 240-248.

Yuxiu, Y. (2024). Application of Translation Technology based on AI in Translation Teaching. Systems and Soft Computing. https://doi.org/10.1016/j.sasc.2024.200072.

Zhang, L. (2023). Empowering Chinese college students in English as a Foreign Language Writing Classes: Translanguaging with translation methods. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1118261.

Zhou, C., & Hou, F. (2025). How Do EFL Teachers Utilize AI Tools in Their Language Teaching? Theory and Practice in Language Studies. https://doi.org/10.17507/tpls.1502.10.