Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

The Impact of AI-Assisted Translation on Arabic-English Translation Quality

Mohammed Kasmi [1], Adil Tajiou[2], Mohamed Boulbali[3]

[1] Applied Communication in Context Laboratory, Mohamed First University, Oujda, Morocco.
[2] Translation Culture and Digital Media Master's Program, Multidisciplinary Faculty, Nador, Morocco.
[3]Translation Culture and Digital Media Master's Program, Multidisciplinary Faculty, Nador, Morocco.

Abstract

The increasing presence of artificial intelligence (AI) in translation practices has raised important questions about its role in translation education, particularly at university level. This study investigates the effects of AI-assisted translation on Arabic–English translation products. Using a pre-test/post-test approach, students were divided into an experimental and control group. They first completed an Arabic–English translation task without AI support, followed by a guided learning session on the responsible use of AI tools for translation. Then, the control group completed a comparable translation task without AI assistance, while the experimental group benefited from the assistance of AI. Translation outputs were assessed using a common evaluation rubric focusing on accuracy, fluency, lexical precision and cultural resonance. In addition, students' experiences and perceptions were collected through a short questionnaire. The findings suggest that AI-assisted translation can support students' learning by improving fluency and lexical choice, while also highlighting the continued need for human intervention to ensure accuracy and contextual adequacy. The study contributes to ongoing discussions on the pedagogical integration of AI in translation education and emphasizes the importance of guided and critical use of AI as a learning tool rather than a substitute for human translation competence.

Keywords: AI-Assisted Translation, Translation Pedagogy, Human-in-the-Loop, Post-Editing.

1. Introduction

Translation pedagogy has traditionally emphasized learner-centered approaches, focusing on the development of linguistic, cultural, and cognitive skills. Kiraly argues that effective translator education positions students as active constructors of knowledge, where problem-solving, reflection, and collaboration are central to learning (Kiraly 12-15). This approach emphasises the importance of human agency and critical thinking, particularly at early stages of translation learning.

Technological advancements have introduced new dimensions to translation education. Bowker reveals that computer-assisted translation tools and machine translation have become essential components of modern translator training (Bowker 27). She emphasizes that technology should not replace critical reasoning but instead serve as a weapon that supports learning and encourages evaluation of machine-generated output. Integrating these tools into pedagogy requires guided approaches that foster reflective engagement with both the source text and the technology itself.

The emergence of AI and digital technologies further challenges traditional assumptions. Cronin contends that the digital revolution has transformed translation into a central paradigm of contemporary culture, suggesting that the "information age" should be termed the "translation age" due to the universal convertibility of digital code and its transformative effects on language, society, and culture (Cronin 3). He highlights the "3T paradigm" of trade, technology, and translation, showing how digital tools shape not only translation practice but also learners' perceptions of what translation entails (Cronin 19). The digital revolution has introduced AI as a transformative force in education and translation. Acquiring the ability to "coexist with Artificial Intelligence" is increasingly essential, since AI can perform tasks requiring human intelligence such as problem-solving and critical thinking (Zouhaier 21). These insights are particularly relevant for educational contexts, where AI-assisted translation can influence students' engagement with meaning, accuracy, and linguistic creativity.

Overall, these works indicate that while AI and digital tools offer valuable support for translation learning through enhancing fluency, lexical choice, and access to authentic language, they also require careful pedagogical framing. Guided and reflective use of these technologies guarantees that students develop not only the ability to produce translations but also the critical thinking skills necessary to evaluate and improve them. The present study builds on this foundation by examining AI-assisted Arabic–English translation among high school students, a context that remains largely unexplored but is increasingly relevant in the digital era.

2. Literature Review

Many studies proved that translation from English into Arabic causes consistent challenges for AI tools. The quality of the Arabic MT translation is still far behind the other languages due to the complex nature of its linguistic system. AI tools are better than CAT tools from English into Arabic and vice versa (Khasawneh & Alsharif, 2025). Another study showed that ChatGPT performed better than Gemini on classical Arabic grammatical literature where it produced a more accurate translation that aligns with the source text in terms of lexical choices accuracy. However, these two models revealed significant limitations, especially when translating culture-specific terms or idiomatic expressions. They specifically struggle with structures that are deeply rooted in the culture or depend heavily on context (Zaimah et al., 2026).

Text typology affects the performance of AI tools in translation. when translating journalistic texts, Reverso, Systran, and Microsoft Azure produced an acceptable translation. However, the major challenges that they faced are the domain-specific terminology and stylistic coherence. This study emphasizes the importance of post-editing and highlights that the human factor in the process of translation is irreplaceable (Bououden, 2025). Similarly, Khoury (2024) compared the performance of Google Neural Machine Translation (GNMT) against human post-editing in the translation of scientific texts from English into Arabic. He concluded that GNMT faced serious challenges when translating scientific terminology and acronyms and that human post-editing remains important for producing a translation that ensures higher level of readability and naturalness. AI has no potential to eliminate the need for human post-editing. Its quality compared to CAT tools or AI is significantly better (Mohammed, 2025; Alsaif et al., 2024).

3. Research Aim and Hypotheses

The purpose of this study was to evaluate the impact of AI-assisted tools on Arabic-to-English translation quality among first year students enrolled in the English department in the Multidisciplinary Faculty in Nador. Using a pretest-post-test control group design with 120 participants, we tested four specific hypotheses.

Hypothesis 1: There will be a statistically significant difference between the scores of the control group on the pretest and post-test.

Hypothesis 2: Experimental Group Improvement: There will be a statistically significant difference between the scores of the experimental group on the pretest and post-test.

Hypothesis 3: There will be no statistically significant difference between the scores of the experimental and control group on the pretest.

Hypothesis 4: There will be a statistically significant difference between the scores of the experimental and control group on the post-test.

4. Method

This study employed a quantitative, pretest-post-test control group design in order to look at how AI can help undergraduate students improve their Arabic–English translation. We used a pre-test/post-test design. First, students translated a short paragraph without any AI help to see their natural performance. Then, they took part in a guided session where they learned how to use AI tools responsibly for translation. After the session, the control group depended on their skills only whereas the experimental group used AI. This allowed us to compare their translations before and after using AI.

5. Research Design

The study unfolded across three phases over a four-week period:

1. **The pre-test phase:**

   Week 1 (Pretest): All participants completed a standardized Arabic-to-English translation task under supervised and AI-free conditions. This translation represented their baseline performance based solely on their linguistic knowledge, intuition, and previously acquired translation skills.

2. **Intervention phase: weeks 2-3**

a. Control group: Received 2 weeks of traditional translation instruction (dictionary use, grammar review, peer feedback, instructor-led error analysis).

b. Experimental group: Received the same core instruction plus a structured module on AI literacy and critical post-editing.

3. **The post-test phase**

   Week 4 (Posttest): Both groups completed a parallel-form translation task. The experimental group was permitted to use the AI tool under the same supervised conditions as the pretest; the control group continued with traditional resources only.

6. Participants

A total of 120 undergraduate students (ages 19–23; M = 20.4, SD = 1.2) enrolled in English studies department in the multidisciplinary faculty in Nador. All participants were native Arabic speakers with intermediate-to-upper-intermediate English proficiency.

Table 1

Group	n	Gender (F/M)	Mean Age
Control	60	34 / 26	20.3 (SD = 1.1)
Experimental	60	36 / 24	20.5 (SD = 1.3)

Preliminary independent-samples t-tests confirmed no significant differences between groups in age, gender distribution, English proficiency ($p > .05$ for all comparisons), supporting the effectiveness of randomization.

7. Research Instruments

Two main research instruments are used in this research:

1. Pretest and post-test process which was conducted over a four-week period. The pretest was taken in week 1 and the post-test was taken in week 4. Students' performance was evaluated based on a translation assessment rubric, which was designed by two certified translation assessors and prepared prior to the study:

table 2: translation assessment rubric

Criterion	Points	Operational Definition	Scoring Anchors
Accuracy	0–6	Faithfulness to source-text meaning; absence of omissions, additions, or distortions	6 = No errors; 4 = Minor deviations; 2 = Major omissions; 0 = Meaning lost
Fluency	0–5	Grammaticality, syntactic naturalness, cohesion, and readability in English	5 = fluent; 3 = Occasional awkwardness; 1 = Frequent errors impede comprehension
Lexical precision	0–5	Precision, register appropriateness, and collocation accuracy of vocabulary	5 = Precise, idiomatic, context-sensitive; 3 = Generally appropriate with minor mismatches; 1 = Frequent poor word choices
Cultural resonance	0–4	Effective handling of idioms, culturally-bound references, and pragmatic nuance	4 = Culturally resonant ; 2 = Literal translation with minor loss; 0 = Misinterpretation of cultural content

2. A structured questionnaire was used as a quantitative research instrument. It aimed to collect statistical data on students' use of AI translation tools, frequency of use, perceived usefulness, and attitudes toward AI-assisted translation. The questionnaire consisted primarily of closed-ended questions (multiple-choice and Likert-scale items), allowing for statistical analysis of trends and patterns. The questionnaire was administered after the translation tasks to ensure that students' responses were informed by direct experience with both human-only and AI-assisted translation processes.

8. Results of the pre-test and post-test

Results related to the first hypothesis

The first hypothesis stated that there would be no statistically significant difference between the experimental and control groups on the pretest. The descriptive statistics in Table 3 show that the two groups began the study with remarkably similar performance. The control group averaged 11.24 (SD = 2.81) on the total score, while the experimental group averaged 10.93 (SD = 3.04). An independent-samples t-test conducted on pretest scores confirmed that this small difference was not statistically significant, ($t(118) = 0.59$, $p = .56$, $d = 0.11$). This non-significant result supports the first hypothesis.

Table 3: Descriptive Statistics of the pre-test and post-test

Criterion	Group	Pretest M (SD)	Posttest M (SD)	Gain Score M (SD)
Total Score (/20)	Control	11.24 (2.81)	13.08 (3.12)	+1.84 (2.19)
	Experimental	10.93 (3.04)	16.42 (2.48)	+5.49 (2.37)
Accuracy (/6)	Control	3.12 (1.14)	3.78 (1.21)	+0.66 (0.98)
	Experimental	3.01 (1.29)	4.52 (1.03)	+1.51 (1.12)
Fluency (/5)	Control	2.83 (1.02)	3.21 (1.09)	+0.38 (0.87)
	Experimental	2.74 (1.18)	4.08 (0.94)	+1.34 (1.01)
Lexical Choice (/5)	Control	2.91 (1.23)	3.38 (1.31)	+0.47 (1.04)
	Experimental	2.82 (1.41)	4.59 (1.08)	+1.77 (1.29)
Cultural Adaptation (/4)	Control	2.38 (0.97)	2.71 (1.08)	+0.33 (0.79)
	Experimental	2.36 (1.12)	3.23 (0.99)	+0.87 (0.93)

9. Results for the second hypothesis:

The second hypothesis predicted a statistically significant difference between groups on the post-test. This is the central test of the intervention's efficacy: if AI-assisted training adds value, the experimental group should outperform the control group after the two-week instructional period. The results strongly support this expectation. As shown in Table 3, the experimental group achieved a mean post-test score of 16.42 (SD = 2.48), compared to 13.08 (SD = 3.12) for the control group—a difference of 3.34 points on a 20-

point scale. The independent-samples t-test on post-test scores in table 2 yielded $t(118) = 8.91$, $p < .001$, with a large effect size ($d = 1.63$). This indicates that the observed difference is highly unlikely to be due to chance and represents a substantively meaningful advantage for the AI-assisted group. This means that the second hypothesis was supported

Table 4: Between-Group Independent Samples t-Tests on Gain Scores

Criterion	t(df)	p-value	Cohen's d	Interpretation
Total Score	t(118) = 8.91	< .001***	1.63	Very large advantage for the experimental group
Accuracy	t(118) = 5.12	< .001***	0.93	Large advantage
Fluency	t(118) = 5.68	< .001***	1.04	Large advantage
Lexical Choice	t(118) = 6.84	< .001***	1.25	Very large advantage
Cultural Adaptation	t(118) = 4.93	< .001***	0.90	Large advantage

10. Results for the third hypothesis:

The third hypothesis anticipated that the control group would show significant improvement from pretest to post-test. This aims to test whether traditional instruction alone produces learning gains, which stands as a necessary benchmark against which to evaluate the added value of AI assistance.

The paired-samples t-test results in Table 4 confirm this expectation. The control group improved from a pretest average of 11.24 to a post-test average of 13.08, a gain of 1.84 points. This improvement was statistically significant, $t(59) = 6.72$, $p < .001$, with a large effect size ($d = 0.87$). In other words, conventional translation pedagogy—practice, feedback, dictionary use—does work. Students learned, and their translation quality improved in measurable ways. Therefore, hypothesis three is clearly supported.

Table 5: Within-Group Paired Samples t-Tests (Pretest vs. Posttest)

Criterion	Group	t(df)	p-value	Cohen's d	Interpretation
Total Score	Control	t(59) = 6.72	< .001***	0.87	Large improvement
	Experimental	t(59) = 18.34	< .001***	2.37	Very large improvement

Criterion	Group	t(df)	p-value	Cohen's d	Interpretation
Accuracy	Control	t(59) = 5.21	< .001***	0.67	Medium-large
	Experimental	t(59) = 10.48	< .001***	1.35	Very large
Fluency	Control	t(59) = 3.38	.001**	0.44	Small-medium
	Experimental	t(59) = 10.31	< .001***	1.33	Very large
Lexical Choice	Control	t(59) = 3.51	.001**	0.45	Small-medium
	Experimental	t(59) = 10.69	< .001***	1.38	Very large
Cultural Adaptation	Control	t(59) = 3.24	.002**	0.42	Small-medium
	Experimental	t(59) = 7.28	< .001***	0.94	Large

11. Results for the fourth hypothesis:

Our fourth and final hypothesis predicted that the experimental group would also show significant pretest-to-posttest improvement—and, implicitly, that this improvement would be substantial given the added support of AI tools. The results not only support this hypothesis but exceed expectations as shown in tables 3 and 5. The experimental group's mean score increased from 10.93 at pretest to 16.42 at posttest, a gain of 5.49 points. The paired-samples t-test yielded $t(59) = 18.34$, $p < .001$, with an exceptionally large effect size ($d = 2.37$). based on this, the fourth hypothesis is supported.

12. Discussion

This study sought to assess the effects of AI-assisted translation in what Cronin terms the "translation age," examining how guided classroom use influences students' awareness of translation quality. The findings indicate that the classroom session played an important role in helping students understand the limits of AI's reliability. Rather than viewing it as a fully automatic solution, students came to recognize AI as a supportive but imperfect tool, a perspective that aligns perfectly with Bowker's idea of technology as a scaffold that supports, rather than replaces, human reasoning.

Through the pre-translation discussion, the translation task, and the post-editing phase, students developed greater awareness of the role of human judgment in translation. Although AI tools were perceived as helpful for improving comprehension and fluency, the comparison between human-only and AI-assisted translations revealed the persistent weaknesses that Bowker cautions against. AI struggled with literal translation and cultural inaccuracy, particularly with terms such as الحناء, confirming that contextual interpretation remains a critical human skill.

The post-editing phase emerged as a central pedagogical component, transforming the task into the reflective process that Kiraly advocates for. The fact that all students modified the AI translations through

collective revision demonstrates that they were not passive recipients but "active constructors of knowledge." Moreover, as students' ethical perceptions evolved to view AI as a neutral learning tool, they demonstrated an emerging ability to "coexist with Artificial Intelligence," a skill Zouhaier identifies as essential for the modern era. Overall, the findings confirm that AI-assisted translation can enhance learning outcomes when integrated critically and supported by systematic post-editing and teacher guidance.

These findings have several practical implications. First, they support the integration of AI literacy and post-editing skills into translation curricula—not as a replacement for foundational instruction, but as a complementary scaffold. Second, they highlight the importance of teaching students to critically evaluate AI output, particularly in domains like cultural adaptation where AI remains less reliable. Third, they suggest that AI tools may help compress the learning curve, enabling students to reach higher performance levels in less time.

13. Limitations and Future Directions

While the results are robust, several limitations need to be considered. First, the study measured immediate post-intervention gains; future research should examine whether these advantages persist over time and transfer to unassisted translation tasks. Second, the findings reflect certain AI tools and one language pair (Arabic-to-English); replication with other tools and language combinations would strengthen generalizability. Third, while the rubric captured key dimensions of translation quality, future studies might incorporate other measures to better understand how students interact with AI suggestions during translation.

14. Conclusion

This study set out to test four clear, methodologically grounded hypotheses about the impact of AI assistance on translation learning. The statistical evidence supports all four: groups were equivalent at baseline, diverged significantly after intervention, and both improved—with the AI-assisted group showing dramatically greater gains. These results do not suggest that AI replaces the translator; rather, they reveal how critically-mediated AI such as ChatGPT and Google Translate can elevate the translator's developmental trajectory, including improving fluency and general comprehension. Nonetheless, they, at times, fall short in handling culturally specific and idiomatic language, often producing literal and contextually inappropriate translations. For educators, the research demonstrates can invested as a supportive tool with human oversight rather than seeing it as a form of cheating. This means that the pending issue is not whether to use AI, but how to use it wisely.

This study shows that the successful integration of AI into translation education does not lie in replacing the human translator, but in fostering a "human-in-the-loop" model. Through guided instruction and reflective practice, students can learn to use AI as a powerful assistant, enhancing their workflow while simultaneously improving their critical thinking and cultural awareness. The research confirms that the future of translation pedagogy is not a choice between human and machine, but a collaborative synergy between the two.

References

1. Alsaif, H. S., & Aluthman, E. S. (2024). An assessment of the quality of post-edited text from CAT tools compared to conventional human translation. Journal of Language Teaching and Research, 15(1), 123-134. doi.org
2. Bououden, R. (2025). Evaluating Arabic-English neural machine translation: Challenges across different text types. Journal of Language and Translation, University of Chlef, 9(1).
3. Bowker, L. (2002). Computer-aided translation technology: A practical introduction. University of Ottawa Press.
4. Cronin, M. (2013). Translation in the digital age. Routledge.
5. Khasawneh, R. R., Alsharif, B. B., & Khasawneh, R. (2025). A comparative study of AI-powered tools for Arabic-English and English-Arabic translation. Journal of Language Teaching and Research, 16(6), 2025-2035.
6. Khoury, O. (2024). Reflection of explicitation in scientific translation: Neural machine translation vs. human post-editing. Journal of Language Teaching and Research, 15(5), 1510-1517.
7. Kiraly, D. (2014). A social constructivist approach to translator education: Empowerment from theory to practice. Routledge.
8. Mohammed, A. S., Ghassemiazghandi, M., & Jamal, M. (2023). Post-editing of neural machine translation of the novel "Murder of the Bookseller" from Arabic into English. International Journal of Language, Literature and Translation, 6(2), 102–114.
9. Mohammed, T. A. S. (2025). Evaluating translation quality: A qualitative and quantitative assessment of machine and LLM-driven Arabic–English translations. Information, 16(6)
10. Slimi, Z. (2023). The impact of artificial intelligence on higher education: An empirical study. European Journal of Educational Sciences, 10(1), 1-15.
11. Zaimah, N. R., Rizqiyah, C., Hadi, S., Muthiah, R., & Nurrohmah Putri, W. (2026). ChatGPT vs Gemini: Which digs deeper into Arabic semantics? Mantiqutayr: Journal of Arabic Studies, 6(1).