International Journal of Linguistics and Translation Studies
Volume 7, Issue1, 2026
Homepage: http://ijlts.org/index.php/ijlts/index
DOI: https://doi.org/10.36892/ijlts.v7i1.666

Exploring the Effectiveness of Pedagogical Orientation of Generative AI Models on Enhancing University Students' Translation Skills: An Experimental Study

Redhwan Alsharef
Department of English language, Faculty of Education & Humanities, Hajjah University, Yemen (Corresponding author)
radhwanalsharif@gmail.com
Faisal Aldawli
Community Center for Translation & Language Teaching– Sana'a University, Yemen
aldfaisal@gmail.com
Khaled Alahdal
Department of Foreign Languages, the University of Nizwa, Sultanate of Oman
khaledalahdal@unizwa.edu.om
Nuha Shamsan
Faculty of Education & Humanities, Hajjah University, Yemen
nuha.ali28@gmail.com

How to cite:
Alsharef, R., Aldawli, F., Alahdal, K., & Shamsan, N. (2026). Exploring the effectiveness of pedagogical orientation of generative AI models on enhancing university students' translation skills: An experimental study. *International Journal of Linguistics and Translation Studies, 7*(1).1–30. https://doi.org/10.36892/ijlts.v7i1.666

ARTICLE HISTORY
Received: *28/10/2025*
Accepted: *20/12/2025*

Keywords:
Translation training; Generative AI; AI-driven Pedagogy; Translation Skills; Intervention; Arab Context

Abstract

This study attempted to bridge the research gap in AI-driven pedagogy for translation training in the Arab context, focusing on the potential of generative AI models to improve the translation proficiency of university translation majors. The research explored the effectiveness of pedagogically oriented generative AI tools in enhancing students' skills across linguistic, cultural, and text-level dimensions in English ⇄ Arabic translation, using a true experimental pre-test-post-test control group design. While both groups used identical training materials, the experimental group received AI-guided training, and the control group was taught through traditional instruction. Through a random sampling (n = 37 per group), participants were recruited from four universities in Yemen and Oman, ensuring a comparable educational background. The findings revealed that the experimental group outperformed the control group in translation achievement in all targeted translation skills due to the impact of guided integration of AI. The study underscored the multi-faceted pedagogical applications of AI in translation education when grounded in a systematic pedagogical framework under instructor guidance. Through highlighting practical pedagogical implications and offering an evidence-based framework for integrating AI into translation programs, the research opens new avenues for innovative practices in AI-assisted translation pedagogy for instructors and curriculum designers.

1. INTRODUCTION

In recent years, artificial intelligence technologies have significantly transformed several fields, including the educational sector and its practices. Despite the widespread application of

AI-powered translation technologies within translation programs, there remains a notable lack of pedagogical foundations for their integration as well as a weak alignment between AI applications and academic instructional approaches (Kanglang, 2021). The existing literature underscores the pressing need for a practical and comprehensive integration of generative AI models in translation education based on pedagogical foundations for enhancing students' translation proficiency (Bakhov et al., 2024; Alghamdi & Alotaibi, 2025; Levin et al., 2025). Further, the existing scholarship emphasizes the need for further exploration of principled approaches to apply AI in translation teaching within higher education institutions (Kanglang, 2021), and addressing how best to incorporate AI-enhanced tools into classroom translation practices and training programs (Kornacki & Pietrzak, 2024). Moreover, there is a critical need to balance AI automation with human insight and supervision through strategic implementation, effective design and training, which can contribute meaningfully to improving modern translation pedagogy (Alghamdi & Alotaibi, 2025; Sadiq, 2025).

Therefore, the current study aims to address these needs and bridge the research gap in the Arab context by examining the effectiveness of a practical, pedagogically-oriented integration of generative AI (ChatGPT-4o & Gemini2.5) in enhancing university students' translation skills and proficiency.

1.1.Research Questions and Hypotheses

1.1.1 Research Questions

The current study aims to answer the following questions:

1- How effective is pedagogically oriented AI training in improving the overall translation skills of university students majoring in English ⇄ Arabic translation?

2- To what extent does structured AI-based training enhance students' translation proficiency in linguistic accuracy, cultural competence, and text-level quality revision and evaluation in English ⇄ Arabic translation?

3- What are the practical pedagogical implications of integrating Generative AI tools in translation education?

1.1.2 Research Hypotheses:

The research has the following two hypotheses and their sub-hypotheses to test:

H01: There is no statistically significant difference in overall translation skills scores between the experimental group and the control group in the post-test.

- H01a: There is no statistically significant difference in linguistic accuracy scores between the experimental group and the control group.
- H01b: There is no statistically significant difference in cultural competence scores between the experimental group and the control group.
- H01c: There is no statistically significant difference in text quality revision and evaluation scores between the experimental group and the control group.

H02: The experimental group shows no statistically significant difference in translation skills performance between pre-test and post-test.

- Ho2a: The experimental group shows no statistically significant difference in linguistic accuracy scores between pre-test and post-test after AI-assisted training.
- Ho2b: The experimental group shows no statistically significant difference in cultural competence scores between pre-test to post-test after AI-assisted training.
- Ho2c: The experimental group shows no statistically significant difference in text-level quality revision and evaluation scores between the pre-test and post-test after AI-assisted training.

1.2. Significance of the Research

Exploring the effectiveness of generative AI tools like ChatGPT and Gemini in translation education is crucial, especially within the Arab context, where this area remains largely unexplored. Addressing this research gap is pressing, particularly with the increasing demand for skilled translators in an AI-driven translation industry (Abu-Rayyash, 2017). This research is intended to demonstrate the practical applications of integrating these tools by positioning generative AI models (ChatGPT & Gemini) as effective pedagogical tools for personalized learning to effectively improve student translation proficiency. Such applications can also redefine instructors' crucial role in AI-assisted translation training, offering a practical guide for educators and curriculum designers in the effective and systematic integration of AI in translation pedagogy. In addition, this research is significant for enhancing pedagogical approaches and preparing students for the evolving landscape of translation competencies, where AI literacy and mastery are essential. Finally, this study contributes a theoretical and practical framework for integrating AI into translation training, providing a pedagogical guide for real-world classrooms and laying the foundation for a free, accessible online training platform based on an AI-driven pedagogical model.

2. LITERATURE REVIEW

2.1 The Evolving Roles of AI Technologies in Translation Education

Research on integrating AI in translation training has progressed from early studies of neural and machine-assisted systems to current interest in generative AI models and their pedagogical applications. Studies consistently find that AI-powered tools facilitate active, personalized, and interactive learning, allowing educators to tailor instruction to diverse student needs (Wang, 2023). AI translation and training systems enhance students' understanding of linguistic structures and provide individualized diagnostic feedback (Brown, 2019; Somers, 2021). Furthermore, platforms that integrate tools like ChatGPT have been found useful in improving student engagement and problem-solving in translation tasks (Hellmich & Vinall, 2021; Xu et al., 2024).

Experimental applications reinforced these benefits, though often with some limitations. For example, Aleedy et al. (2022) developed a deep-learning chatbot for Arabic translation feedback, though it was not based on large language models and was confined to sentence-level correction. Bakhov et al. (2024) found that an AI-assisted application improved students' translation quality and motivation in a Ukrainian university course with a module focused on

translating specific linguistic items. Xu (2024) advocated for AI empowerment in undergraduate translation education, showing how intelligent teaching platforms and personalized learning systems enhance both pedagogy and teacher professional development. Through supporting assessment, generating feedback, and identifying learning difficulties, generative AI tools such as ChatGPT and Gemini illustrate their direct utility in advancing pedagogical practices (Grassini, 2023; Kasneci et al., 2023). Alghamdi and Alotaibi (2025) demonstrated ChatGPT-4o's reliability for assessment in legal translation, advocating for its strategic pedagogical implementation under human instructor's guidance.

The literature also indicates that effective AI integration requires re-skilled educators, redesigned curricula, and collaborative frameworks between translation and technology specialists (Wang, 2023; Koka, 2024; Khasawneh & Shawaqfeh, 2024), along with constant professional development to adapt to evolving technologies (Alharbi, 2024). Further, Sadiq (2025) highlighted that detailed prompts can enhance AI performance in translation assessment, calling for curricula that address tool strengths and limitations.

2.2 Challenges in AI Integration in Translation Education

Despite AI potential, significant challenges persist such as concerns about ethical use, data privacy, and professional replacement of human instructors (Grassini, 2023). Students and educators express their concerns about over-reliance on AI, fearing it may lead to diminishing skills even as they value its efficiency (Amaro & Pires, 2024; Atlas, 2023). Technically, AI tools still struggle with textual coherence, cohesion, and cultural nuances (Xu et al., 2024), while issues of accessibility, usability, and transparent assessment frameworks further complicate implementation (Koka, 2024; Khasawneh & Shawaqfeh, 2024).

These challenges reflect a consensus that AI must act as a complementary tool to, not a replacement for, human expertise. In response, scholars advocate for balanced, hybrid models that integrate human critical insight with machine efficiency (Cheng, 2022). However, achieving this integration requires institutional support for curriculum reform, instructor training, and empirical research grounded in pedagogical theory (Omar & Salih, 2024).

2.3 From Conceptual Advocacy to Experimental Testing

Some experimental and mixed-methods investigations have begun to move beyond descriptive accounts toward testing AI efficacy in translation education. Emara (2024) conducted a quasi-experimental study comparing the effectiveness of NMT tools (Google Translate, Reverso) and LLMs (ChatGPT, QuillBot) in teaching translation skills. Although the study was not pedagogically-oriented, it revealed a positive impact of LLMs on improving students' translation skills. Similarly, Ed-Dali (2025) compared DeepSeek R1 and ChatGPT- 4.5 in Arabic-English literary translation, proposing a hybrid approach where AI functions as a scaffolding tool dependent on human post-editing. Moreover, Alghamdi and Alotaibi (2025) empirically investigated ChatGPT reliability for assessment in legal translation, emphasizing its strategic pedagogical implementation under human instructor's guidance. Sadiq's (2025) comparison of AI and human translator output quality found that detailed and specific prompts can significantly improve AI assessment performance. Accordingly, his study recommended developing curricula that address both the strengths and limitations of AI.

At the conceptual level, Hajdu and Farkas (2025) argued for balanced AI integration, emphasizing the enduring role of human linguistic judgment, the necessity of critical post-editing, and the foundational importance of AI literacy for effective training programs. Previous scholarship has similarly pinpointed significant untested areas. For instance, Mohammed and Aljanabi (2024) conceptualized a framework for real-time AI-assisted translation quality assessment, though their study lacked empirical evaluation of learning outcomes. Mohsen (2024) compared the translation accuracy of LLMs and Google Translate but did not extend the analysis to pedagogical implementation. Meanwhile, Tian (2024) proposed an adaptive AI-based model for translator training, and Al-Ali (2025) surveyed student perceptions of AI's impact on skill development.

This conceptual work is supported by relevant theoretical scholarship advocating for AI literacy and a constructionist reframing of education (Kiraly, 2014; Zhang & Doherty, 2025; Levin et al., 2025). Overall, the current literature reveals strong theoretical interest in generative AI for translation education but shows limited experimental verification of its comprehensive pedagogical potential.

2.4 Research Gap in the Arab Context and the Rationale for the Present Study

While existing research has confirmed AI utility for discrete translation tasks and advocated for its pedagogical integration, a critical gap remains between theoretical potential and practical classroom implementation, particularly within the Arab context. Current studies have largely focused on evaluating AI output quality and its reliability in translation assessment (Alghamdi & Alotaibi, 2025; Mohammed, 2025; Sadiq, 2025), or comparing generative AI with NMT, rather than empirically examining the AI pedagogy integration for human skill development. Even where skills were addressed, such as in the study by Emara (2024), training remained product-oriented rather than pedagogically grounded. Moreover, most evidence stems from East Asian and Western contexts, leaving the Arab higher-education environment critically under-researched (Omar & Salih, 2024). Consequently, there is a clear absence of experimentation regarding how generative models like ChatGPT or Gemini can be pedagogically oriented to enhance English-Arabic translation skills of university students.

Accordingly, the current study addresses this research gap by implementing a guided pedagogical intervention that transfers constructionist theory (Kiraly, 2014; Levin et al., 2025) into a practical model for Arabic-English translation skills. Through this pedagogical framework, the research aims to shift the focus in translation training from product to process, positioning generative AI as a multi-faceted pedagogical agent under instructor supervision to foster students' learning experience and critical skills. By empirically validating this approach, the research seeks to provide an evidence-based blueprint for human-AI collaboration in translator education, capable of yielding transformative skill improvements for university students.

3. THEORETICAL FRAMEWORK

While pedagogy traditionally involves curriculum design, instructional delivery, and assessment (Richards & Schmidt, 2002), the rapid integration of generative AI imposes a paradigm shift in translation education. Moving beyond traditional transmission models, where knowledge is passively transmitted, this study adopts Constructivism as its primary theoretical framework. This choice is based on the theory direct alignment with the interactive, process-oriented nature of the study developed AI-assisted training modules.

Constructivism asserts that learners should not passively receive information, but rather actively build knowledge through experience, collaboration, and contextualized practice (Li, 2014). In the specific context of translation studies, Kiraly (2014) argues for a shift from the teacher as a sole source of knowledge to a facilitator who guides learners through "emergent translation competence." This approach emphasizes that translation skills are not merely taught but are constructed through active inquiry and problem-solving (Elen et al., 2007; Kalpana, 2014).

The current study integrates generative AI models like ChatGPT and Gemini to operationalize constructivist principles and create a student-centred environment. While constructivism has traditionally emphasized human-to-human interaction, Levin et al. (2025) posit that, from this perspective, generative AI transcends the role of a mere automation tool to become an active "partner to think with." Accordingly, in this framework, the AI functions as a cognitive scaffold or mediator, providing adaptive, real-time feedback. This interaction enables learners to co-construct meaning by engaging with the generative AI tool to navigate complex translation nuances, aligning with Kiraly's (2014) model of translation as a situated, collaborative process.

Furthermore, the study aligns with the constructionist aspect of learning (learning by making) where students actively generate, critique, and refine translation artifacts (Levin et al., 2025). This process relies heavily on ‘reflection’, defined as the process of looking back on learning experiences to understand their significance (Richards & Schmidt, 2002). Operationalized through the critical analysis of AI-generated outputs, reflection ensures that students do not passively accept AI suggestions. As Levin et al. (2025) note, this new dynamic necessitates careful critical analysis to navigate the shift effectively. By reflecting, criticizing, and refining AI outputs, students transform the translation process into a personalized experience where they actively construct their own evaluative skills and translation strategies. Therefore, constructivism serves as the robust foundation for this study, aligning the theoretical ideals of active, situated learning with the practical application of AI-based pedagogy training.

4. METHODOLOGY

4.1.Research Design

The study adopted a true experimental, pre-test-post-test control group design with experimental and control groups. This design is highly effective for establishing causality, as it measures the impact of an intervention while controlling for extraneous variables (Shadish et al., 2002; Campbell & Stanley, 2015). Its primary strength lies in random assignment, which creates statistically equivalent groups before the intervention to minimize selection bias and strengthen the validity of any causal inferences (Bhattacherjee, 2019; Creswell & Creswell, 2017). The independent variable of the study was the type of instruction: AI-assisted training versus traditional instruction. The dependent variables were 1) overall translation performance and 2) specific sub-skills scores in linguistic accuracy, cultural competence, and text-quality revision and evaluation. A quantitative approach was used to analyse pre-test and post-test results. For gaining deeper pedagogical insights and addressing the third research question, a post-training survey was administered to the experimental group to gather their perceptions and provide context for interpreting the quantitative findings.

4.2. Participants and Sampling

The target population comprised third-year undergraduate translation majors at four universities: Sana'a University, the Yemeni Jordanian University, and Al-Nasser University in Yemen, and the University of Nizwa in Oman. A random sample was drawn from institutional registries of eligible students during the 2025–2026 academic year. The inclusion of an Omani university was a strategic decision to broaden the Arab contextual scope of the research, facilitated logistically by a team member at that institution.

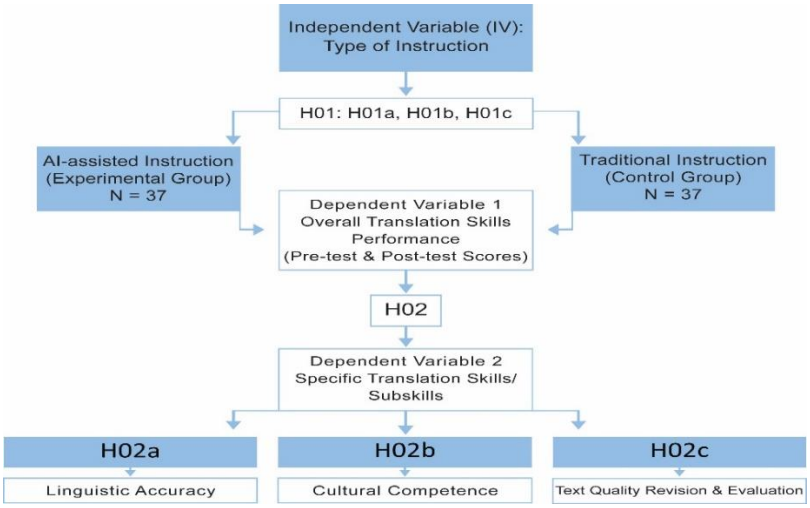
Participants were 74 third-year translation majors, randomly sampled from the four universities in Yemen and Oman. From this pool, volunteers provided informed consent. These consented individuals were then randomly assigned to an experimental group (n=37) or a control group (n=37) using a random number table. All participants met the following inclusion criteria: (1) enrollment as a third-year translation major, (2) intermediate English proficiency, (3) completion of introductory translation courses, and (4) no prior formal training in AI-based translation tools.

Both groups completed four training sessions (two hours each session) over two weeks, covering identical core content, differing only in intervention type (pedagogical approach). The experimental group received guided training using ChatGPT-4o and Gemini 2.5. The control group completed the same tasks via traditional, instructor-led methods without AI. Two trained instructors followed a standardized protocol to ensure consistency.

To ensure consistency and intervention fidelity, two trained instructors (members of the research team) delivered the sessions. Both instructors followed a detailed, standardized instructional guide and participated in standardisation meetings prior to the study. Figure 1 summarizes relevant details about the study experimental intervention, variables and sample size.

Figure 1

Experimental research structure & variables



4.3.Research Instruments and Training Materials

4.3.1 Translation Skills Tests (Pre-test/Post-test)

To measure the effect of the training intervention, the same pre-test and post-test instruments were developed for both experimental and control groups. These tests quantitatively assessed participants' translation performance across three targeted skill domains: linguistic accuracy, cultural competence, and text quality revision and evaluation, establishing a baseline for comparison before and after the intervention.

The pre-test and post-test had identical content, questions and structure; each test contained 30 items (25 multiple-choice, 5 open-ended) with a total score of 60. To ensure reliability and prevent answer memorization, identical linguistic and cultural concepts were tested across both tests, but presented in different contextual activities and sentences. An answer key facilitated the scoring of objective items, while a validated rubric was used for scoring the open-ended questions. Click here to see the pre/post-test.

4.3.2 Post-training Survey

A post-training questionnaire was administered only to the experimental group (N=37) to capture qualitative insights into the pedagogical potential of the generative AI tools. The survey consisted of 13 items on a 5-point Likert scale (ranging from Strongly Disagree to Strongly Agree), organized into two subscales: AI Impact on Translation Skills (6 items) and Overall Pedagogical Benefits (7 items). Clink here for more details about the post-training survey.

4.3.3 Validity and Reliability of Research Instruments

Instruments validity was established through expert review. A panel of six university experts in translation studies and applied linguistics evaluated the content validity of the tests, questionnaire and training modules using the Content Validity Index (CVI), meeting the established thresholds (I-CVI $\geq$ 0.78; S-CVI/Ave $\geq$ 0.90) (Polit & Beck, 2006). All instruments exceeded established thresholds: the questionnaire achieved an S-CVI/Ave of 0.954, the training modules yielded an S-CVI/Ave of 0.946, and the pre-test and post-test scored an S-CVI/Ave of 0.938, revealing high validity. Further, the validators' qualitative feedback in the instruments' open-ended items led to some minor refinements in wording and some items clarity before implementation.

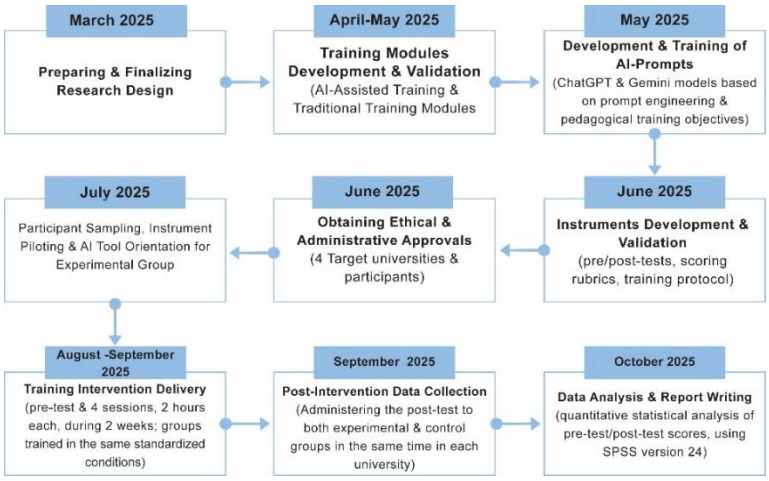
The pre-test and post-test reliability was ensured through pilot testing with 30 students (similar to the target participants), confirming test clarity and appropriate difficulty. This led to some necessary adjustments in some questions phrasing and test duration. Regarding inter-rater reliability for tests open-ended items, four raters independently scored responses using a detailed rubric, resulting in an ICC of 0.85, indicating good reliability. Similarly, a pilot of the AI-assisted training modules with 10 students led to practical improvements, including rewording some activity instructions, prompts and adjusting sessions duration. Moreover, the reliability of the questionnaire was evaluated using Cronbach's alpha, showing good internal consistency for the overall item scale (α = .806), and within the two subscales: AI Impact on Translation Skills (α = .704) and Broader Pedagogical Benefits (α = .803).

4.4. Study Procedures

The study was conducted in sequential phases and procedures to ensure methodological rigor for the experimental intervention, as outlined in Figure 2.

Figure 2

Visualized Timeline & Procedures of the Study Intervention



4.4.1 Preparation Phase

The preparatory phase involved several procedures like finalizing research instruments and the AI-assisted training materials, followed by expert validation and obtaining formal ethical approvals. Other procedures included participant meetings, sampling, and collecting informed consent, concluding with a pilot test of the instruments.

4.4.2 Pre-test

The researchers administered a pre-test to consenting participants (n=37 per group) at four universities, assessing three areas of translation skills: linguistic accuracy, cultural competence, and text quality revision in alignment with the research questions and training modules. At each university, the pre-test was administered simultaneously for both experimental and control groups. Participant responses were anonymized via a coding system for subsequent scoring.

4.4.3 Orientation Session

Before the intervention, the experimental group received an orientation session in which they were trained on how to appropriately use the generative AI (ChatGPT & Gemini) within AI-assisted training modules and how to manage technical issues during the coming training sessions.

4.4.4 Training Plan for Intervention

A unified training plan was established and implemented by the instructors for both groups to ensure methodological consistency, aligning with the research objectives. Accordingly, the instructors adhered strictly to the training plan: managing session duration, procedures, and content coverage to ensure the intervention validity and prevent confounding effects.

4.4.5 Training Intervention

Both groups received the same training content over four sessions within a two-week period (2 hours for each session), though scheduling varied by university. Instruction was delivered simultaneously by trained instructors following a standardized plan. The experimental group worked with an instructor-facilitator using AI tools and a digital workbook, while the control group received traditional, instructor-led instruction using printed materials.

4.4.6 Post-test

The same post-test was administered to both groups to measure proficiency gains. While matching the pre-test in content, length, and scoring, the post-test item formats were altered to reduce reliance on recall and enhance assessment validity, helping students apply learned structures to new contexts. For example, a pre-test item testing the translation of the preposition collocation structure مشهور أو يشتهر ب "famous for" ("Is Kenya famous for growing tea?") was altered in the post-test to "Sana'a is famous for ancient buildings."

4.4.7 Translation Training Modules

The researchers developed two parallel sets of pedagogical training modules: AI-assisted translation training for the experimental group and traditional instruction for the control group. Both sets targeted the same translation skills: linguistic accuracy, cultural competence, and text-quality revision with identical content and duration. Table 1 shows the same training skills and contents delivered for the two groups.

Table1

Targeted Translation Skills

Linguistic accuracy Domain	**Grammar:** Addressing syntactic differences through re-directional translation context (e.g., prepositions, reported speech, exclamation). **Vocabulary:** Resolving lexical challenges and confusing synonyms through translation context (e.g., borrow/lend, testify/certify).
Cultural Competence Domain	Translating culture-bound expressions including idiomatic expressions and social/religious nuances using appropriate strategies/ approaches. (e.g., God forbids! /Even Homer sometimes nods/ يد الله مع الجماعة/النبي يونس/النبي أيوب).
Text quality revision & Evaluation Domain	Proofreading translated texts for spelling and linguistic errors; evaluating translation quality based on accuracy, fluency, naturalness, cultural adaptation, and style.

The crucial difference between the modules sets was in the instructional method used. While the control group modules used traditional, instructor-led techniques, the experimental group modules were structured around pedagogically engineered prompts designed to use generative AI (ChatGPT-4o & Gemini 2.5) as a tutor, feedback generator, and object of critical analysis. These prompts were integrated into a structured AI-assisted training modules, facilitating a consistent cycle of task completion, documentation, and critical reflection.

4.4.8 Methods for Determining Targeted Translation Skills

The targeted translation skills were determined through triangulated methods, including a review of relevant literature and existing Arab academic translation curricula, analysis of common student errors from teaching experience, and consultation with fellow translation instructors as well as drawing on the researchers' academic teaching experience of translation courses and familiarity with students' common translation challenges. This approach ensured

the selected skills were grounded in both academic standards and the practical needs of third-year translation majors before being operationalized into training activities.

4.4.9 Activity Structure and Prompt Design in the AI-Assisted Training Modules

The AI-assisted training modules were built on a structured pedagogical framework to ensure systematic application. Each module followed a consistent format: clear learning objectives, an instructor-led introduction, pre- and post-activity assessments, and collaborative reflection tasks. Central to this design were meticulously engineered prompts, iteratively refined through piloting to elicit reliable and instructionally aligned outputs from ChatGPT and Gemini.

These pre-designed prompts served specific pedagogical functions, such as structuring clear AI roles, scaffolding comparative analysis between student work and AI feedback, and fostering critical evaluation using techniques like few-shot prompting. This guided prompting was embedded within a larger activity structure that required students to document prompts, AI responses, and their own work in a shared Google Doc workbook. Obligatory reflection prompts (e.g., “What is the most important correction you learned?”) and collaborative discussions then facilitated critical engagement with the AI output. For instance, when asked to identify the translation approach used in an activity (e.g., classifying the Arabic translation of "Never say die" (لا تقنط من رحمة الله) as either Semantic or Communicative), the AI models initially showed inconsistency. Providing simplified definitions of these approaches, adapted from theorists like Newmark (1988), oriented them to apply the correct analytical framework consistently, resulting in more accurate and reliable evaluations for culturally nuanced tasks. For more details about AI-assisted translation training modules, click here.

4.5.Data Analysis

The collected data were analysed to address the research questions and test the corresponding null hypotheses. SPSS (Version 24) was used to analyse quantitative data from the pre-tests and post-tests for both the experimental and control groups. To ensure scoring consistency, objective items (n=25) were evaluated with a verified answer key, while open-ended responses (n=5) were assessed using a validated scoring rubric.

Due to the nature of the study hypotheses, specific statistical tests were selected. For comparing the performance of the two independent groups, the study adopted the Mann-Whitney U test, while the Wilcoxon Signed-Rank test was used to evaluate the progress within the experimental group from pre-test to post-test. The choice of these non-parametric tests was determined by the results of the data normality testing, as detailed in the results section. An experienced statistician conducted the analysis of these non-parametric tests: the Mann-Whitney U test and the Wilcoxon Signed-Rank test.

The qualitative findings from a post-training survey were analysed in order to address the third research question concerning pedagogical implications of generative AI in translation training. The survey was administered only to the experimental group (N=37). Responses were analysed using descriptive statistics (means, standard deviations, percentages using SPSS version 24) to delineate participants’ perceptions of the AI-assisted training potential and effectiveness in translation pedagogy.

5. RESULTS

5.1. The Analysis of the Pre-test and the Post-test

5.1.1. Analysis of Normality Testing

Before hypotheses testing, the assumption of normality was assessed using the Shapiro-Wilk test, as it is recommended for samples of $n \leq 50$ (Pallant, 2020). Based on the standard null hypothesis for normality tests, a non-significant result ($p \geq 0.05$) indicates normality, while a significant result ($p < 0.05$) indicates a violation of normality, necessitating non-parametric alternatives (Field, 2018). The Shapiro-Wilk test results for the current data were significant ($p < 0.05$), violating the assumption of normality. Consequently, non-parametric tests were employed for all subsequent analyses, as illustrated in the following tables and histograms: The results, presented in Table 2, indicated that while pre-test scores and the control group's post-test were normally distributed ($p > .05$), the experimental group's post-test scores for overall translation skills significantly deviated from normality ($p = 0.011$). A detailed analysis of the three sub-skill domains (linguistic accuracy, cultural competence, text-quality revision) further confirmed violations of normality in five variables as shown in Table 3.

Table 2

Normality Test Results of Overall Translation Skills

	Shapiro- Wilk Test			
Variable	Statistics	df	p-value	Normality
Control Group Pre	0.973	37	0.489	Normal
Control Group Post	0.978	37	0.657	Normal
Experimental Group Pre	0.975	37	0.558	Normal
Experimental Group Post	0.920	37	0.011	Not Normal

Consequently, and to ensure the robustness of the analysis, non-parametric tests were employed for all subsequent inferential analyses and hypothesis testing as follows:

- The Mann-Whitney U test was used for independent groups comparisons (experimental vs. control group), addressing hypotheses H01, H01a, H01b, and H01c.
- The Wilcoxon Signed-Rank test was used for within-group comparisons (pre-test vs. post-test within both groups), addressing hypotheses H02, H02a, H02b, and H02c.

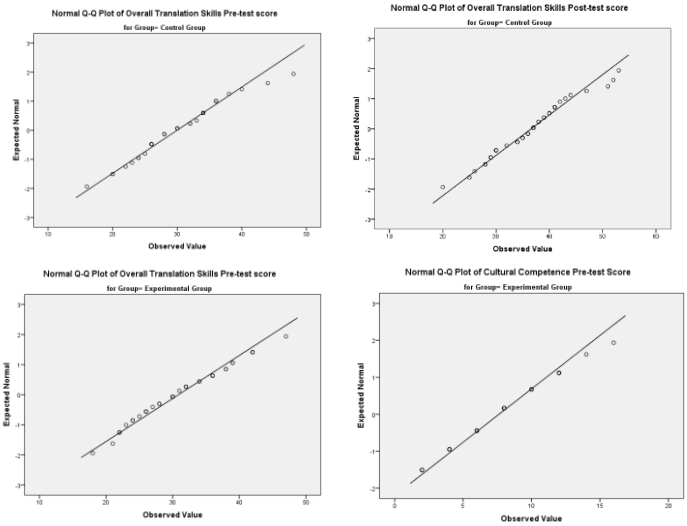
Table 3

Normality Test Results of Specific Translation Skills

Variable	Group	Statistic	df	P-value	Normality
Linguistic Accuracy	Control Group Pre	0.951	37	0.104	Normal
	Experimental Group pre	0.947	37	0.078	Normal
	Control Group Post	0.952	37	0.110	Normal
	Experimental Group Post	0.882	37	0.001	Not normal
Cultural Competence	Control Group Pre	0.937	37	0.038	Not normal
	Experimental Group pre	0.954	37	0.127	Normal
	Control Group Post	0.902	37	0.003	Not normal
	Experimental Group Post	0.862	37	0.000	Not normal
Text quality revision	Control Group Pre	*0.958*	*37*	0.178	Normal
	Experimental Group pre	*0.966*	*37*	0.305	Normal
	Control Group Post	*0.972*	*37*	0.455	Normal
	Experimental Group Post	*0.933*	*37*	0.027	Not normal

Figure 3

The Scatter Plots of the Data Normality



5.1.2. Baseline Equivalence of the Independent Groups

The Mann-Whitney U test was used to assess whether a significant difference existed in the overall translation skills pre-test scores between the experimental and control groups before the intervention to compare their performance at baseline.

Table 4

Mann-Whitney U Test Results of EG & CG's Pre-Test Scores

Variable	Group	N	Mean Rank	Sum of Ranks	U	*p* value
Overall Translation Skills Pre-test score	Control Group	37	36.27	1342	639	0.622
	Experimental Group	37	38.73	1433		

As shown in Table 4, the results indicate no statistically significant difference in overall translation skills between the control group (Mean rank =36.27) and the experimental group (Mean Rank =38.73). The table also shows that there was no statistically significant difference between the scores of the two groups' pre-test (U = 639, p = 0.622/p>0.05). This confirms the two groups were equivalent in their overall translation proficiency prior to the intervention, strengthening the internal validity of the study and confirming that any post-intervention differences could be attributed to the intervention.

5.2. Between-Groups Post-Intervention Comparisons

5.2.1. AI Impact on Overall Translation Skills

The Mann-Whitney U test was used to test the null hypothesis (H01) that there is no statistically significant difference in overall translation skills post-test scores between the experimental and control groups.

Table 5

Between-Groups Comparison: Mann-Whitney U Test for Post-test Scores

Variable	Group	N	Mean Rank	Sum of Ranks	U	*p* value
Overall Translation Skills Post-test score	Control Group	37	21.84	808	105.000	0.001
	Experimental Group	37	53.16	1967		

Total sample size in both groups = 74

The results, presented in Table 5, revealed a statistically significant difference between overall translation skills post-test scores of the experimental group and the control group (U = 105.000,

p<.05). The experimental group achieved a substantially higher mean rank (53.16) compared to the control group's mean rank (21.84), indicating the positive impact of the AI-assisted training on the experimental group's performance. Since the p-value = 0.001 (p<.05), the null hypothesis (H01) is rejected. These results demonstrate that the pedagogically structured AI training was more effective than traditional instruction, yielding significantly higher overall translation achievement by the experimental group.

5.2.2. Mann-Whitney U Test Analysis of Translation Sub-skills

Mann-Whitney U tests were also performed to test sub-hypotheses H01a, H01b, and H01c related to the three sub-skill domains. As summarized in Table 6, statistically significant differences in favour of the experimental group were found for all three domains:

Table 6

Mann-Whitney U Test Results for EG & CG's Translation Sub-skills Post-test

Variable	Group	N	Mean Rank	Sum of Ranks	U	*p* value
Linguistic Accuracy	Control Group Post	37	24.11	892.00	189.000	0.001
	Experimental Group Prost	37	50.89	1883.00		
Cultural competence	Control Group Prost	37	25.68	950.00	247.000	0.001
	Experimental Group Post	37	49.32	1825.00		
Text quality revision & evaluation	Control Group Post	37	22.23	822.50	119.500	0.001
	Experimental Group Post	37	52.77	1952.50		

Note: All p-values are significant at $p < 0.05$.

- Linguistic Accuracy (H01a): The result was significant: ($U = 189.000$, $p<.05$), with the experimental group's higher mean rank (50.89 vs. 24.11 for the CG) demonstrating the superior effectiveness of AI-assisted training on improving linguistic accuracy.
- Cultural Competence (H01b): A statistically significant difference was found: ($U = 247.000$, $p < .005$). The experimental group's higher mean rank (49.32 vs. 25.68) suggests that the pedagogically guided AI tools effectively enhanced skills related to cultural competence.
- Text Quality Revision & Evaluation (H01c): The result revealed a highly significant difference in the two groups' performance ($U = 119.500$, $p < .005$). With the experimental group achieving a much higher mean rank (52.77 vs. 22.23 for CG), the results demonstrate the positive effect of the AI training on enhancing quality revision and evaluation skills.

Consequently, all three sub-hypotheses (H01a, H01b, H01c) are rejected. The experimental group's consistently higher mean ranks demonstrate the broad effectiveness of the AI-assisted training across the targeted skill dimensions. The consistent rejection of all null hypotheses provides robust quantitative evidence for the transformative effectiveness of pedagogical integration of generative AI tools (ChatGPT & Gemini) into translation education.

5.3. Within-Group Analysis: Wilcoxon Signed-Rank Test for the Experimental Group's Post-test

To test H02, the Wilcoxon Signed-Rank Test was used to evaluate the within-group progress of the experimental group after conducting the AI-assisted training intervention. For all Wilcoxon Signed-Rank tests, the alpha level was set at (.05), meaning the null hypothesis would be rejected if $p \leq .05$, leading to the conclusion of a statistically significant difference.

5.3.1. Wilcoxon Signed-Rank Test Analysis of Overall Translation Skills Performance

The Wilcoxon Signed-Rank test results (as presented in Table 7) showed a statistically significant difference in the overall translation skills scores of the experimental group students between pre-test and post-test scores in favour of the post-test ($z = -5.306$, $p < 0.05$), with a very large effect size ($r = 0.87$).

Table 7

Wilcoxon Signed-Rank Test for the Experimental Group's Overall Translation Skills

Item	N	Mean Rank	Sum of Ranks	Z-value	P-value	R
Negative Ranks	0	0	0.00	-5.306	0.001	0.87
Positive Ranks	37	19	703			
Ties	0					
Total	37					

Table 8

Descriptive statistics of EG's pre-test and post-test Scores in Overall Translation Skills Performance

Variable	N	Mean	Median	Standard Deviation
Experimental Before	37	30.86	30.00	7.001
Experimental After	37	50.43	51.00	5.06

As seen in Table 7, all 37 participants in the experimental group showed significant improvement (37 positive ranks, 0 negative ranks) in their overall translation proficiency from pre-test to post-test, indicating the positive impact of the AI pedagogy-driven training. Due to the data deviation from normal distribution, the median score was analysed (Conover, 1999; Thowaini & Qassem, 2024), revealing a substantial increase from (30.00) to (51.00), a gain of 21 points as shown in Table 8. This 21-point median gain, supported by a significant p-value ($p < .001$) and a very large effect size, provides strong evidence to reject the null hypothesis (H_{02}), confirming a statistically significant and substantial enhancement of the experimental group's overall translation skills through AI-assisted training.

5.3.2. Analysis of Wilcoxon Signed-Rank Test for Translation Sub-Skills Performance

The analysis was extended to the specific sub-skills to test the corresponding sub-hypotheses ($H_{02}a$, $H_{02}b$, $H_{02}c$). The results, presented in Table 9, demonstrate statistically significant improvements in the experimental group's post-test across all three domains.

Table 9

Analysis of Wilcoxon Signed-Rank Test Results for EG's Translation Sub-Skills

Sub-skill		N	Mean Rank	Sum of Ranks	Z-value	P-value	R
Linguistic Accuracy Before-After	Negative Ranks	2	4.00	8.00	-5.050[b]	0.001	-0.83
	Positive Ranks	33	18.85	622.00			
	Ties	2					
	Total	37					
Cultural Competence Before-After	Negative Ranks	3	3.00	9.00	-5.124	0.001	-0.84
	Positive Ranks	33	19.91	657.00			
	Ties	1					
	Total	37					
Text Quality Revision Before-After	Negative Ranks	1	1.50	1.50	-5.286	0.001	-0.87
	Positive Ranks	36	19.49	701.50			
	Ties	0					
	Total	37					

Based on negative ranks (Z-value).

Table 10

Descriptive statistics of EG's Pre-test and Post-test Scores in Translation Subskills

Variable	N	Mean	Median	Standard Deviation
Linguistic Accuracy Before	37	11.03	12.00	3.39
Linguistic Accuracy After	37	16.97	18.00	2.24
Cultural Competence Before	37	7.62	8.00	3.46
Cultural Competence After	37	13.51	14.00	2.38
Text Quality Evaluation Before	37	12.97	12.00	3.83
Text Quality Evaluation After	37	19.95	20.00	2.55

Linguistic Accuracy (H02a)

Regarding Linguistic Accuracy ($H_{02}a$), analysis of the experimental group's performance before and after the guided AI translation training revealed a statistically significant improvement. This is evidenced by the Wilcoxon Signed-Rank Test, as indicated in Table 9, which shows a significant difference between pre-test and post-test scores ($Z = -5.050$, $p < .001$) with a very large effect size ($r = 0.83$). The distribution of ranks, with 33 positive ranks against only 2 negative ranks, confirms the improvement for the vast majority of participants. Furthermore, the median score increased substantially from (12.00) in the pre-test to (18.00) in the post-test, as seen in Table 10.

Cultural Competence (H02b)

A statistically significant and substantial improvement was also found for cultural competence. The Wilcoxon Signed-Rank Test indicated a significant difference between pre-test and post-test scores ($Z = -5.124$, $p < .001$) in favour of the post-test with a very large effect size ($r = 0.84$). As Table 10 shows, the median score increased substantially from (8.00) to (14.00) with a 6-point gain. This improvement is indicated by 33 positive ranks against only 3 negative ranks.

Text Quality Revision & Evaluation (H02c)

In Text Quality Revision & Evaluation domain, the results showed a highly statistically significant difference ($Z = -5.286$, $p < .001$) in favour of the post-test, with a very large effect size ($r = 0.87$). As noticed in Table 10, the median score demonstrated a marked increase, rising

from (12.00) in the pre-test to (20.00) in the post-test. Notably, 36 out of 37 participants showed improvement (positive ranks), with only one negative rank.

The Wilcoxon Signed-Rank test results consistently demonstrate that the experimental group made statistically significant progress from pre-test to post-test, with p values less than 0.05 ($p < .05$) and very large effect sizes observed across all three sub-skills. Consequently, all null hypotheses (H02, H02a, H02b, and H02c) are rejected, providing robust evidence for the effectiveness of the pedagogically guided AI-training in enhancing translation performance in general and in specific skills in particular.

5.4. Comparison of the Effect of the Intervention Type

Table 11 highlights the divergent outcomes attributed to the intervention type. While both groups began at comparable baselines in overall translation skills and sub-skills, their post-intervention outcomes differed significantly. While the control group made modest median gains in overall translation performance (+7 points), the experimental group's gains were markedly larger (+21 points). This progress is consistently observed at the sub-skill level: the experimental group showed higher learning gain in linguistic accuracy and cultural competence (+6 points) than the gain achieved by the control group (+ 2 points). The most prominent difference was in text quality revision and evaluation, where the experimental group's 8-point gain vastly exceeded the control group's 2-point gain. These results underscore the transformative effectiveness of the AI-driven pedagogy, demonstrating its clear advantage over traditional instructional methods for developing translation skills.

Table 11

Comparing the Median Scores of CG & EG Before and After Intervention

Skill / Variable	Group	Pre-test Median	Post-test Median	Positive Median Difference
Overall Translation Skills	Control Group (CG)	30	37	+7
	Experimental Group (EG)	30	51	+21
Linguistic Accuracy	Control Group (CG)	10	12	+2
	Experimental Group (EG)	12	18	+6
Cultural Competence	Control Group (CG)	8	10	+2
	Experimental Group (EG)	8	14	+6
Text Quality Revision & Evaluation	Control Group (CG)	13	15	+2
	Experimental Group (EG)	12	20	+8

5.5.Analysis of Post-Training Questionnaire: Perceived Effectiveness and Pedagogical Implications of AI-assisted Translation Training

A post-intervention questionnaire was administered to the experimental group (N=37) to explore their perceptions on the pedagogical effectiveness and usefulness of Generative AI tools (ChatGPT & Gemini) in translation training. The purpose of this post-training survey was to gain more in-depth insights from the students on the training intervention and the practical pedagogical implications of using AI in translation education, as well as to support the quantitative findings collected through the tests. This post-training survey adopted a 5-point Likert scale (Strongly Disagree – Disagree – Neutral – Agree – Strongly Agree). The questionnaire results, summarized in Tables 12 and 13 and Figures 4 and 5, indicate student perceptions across two major dimensions.

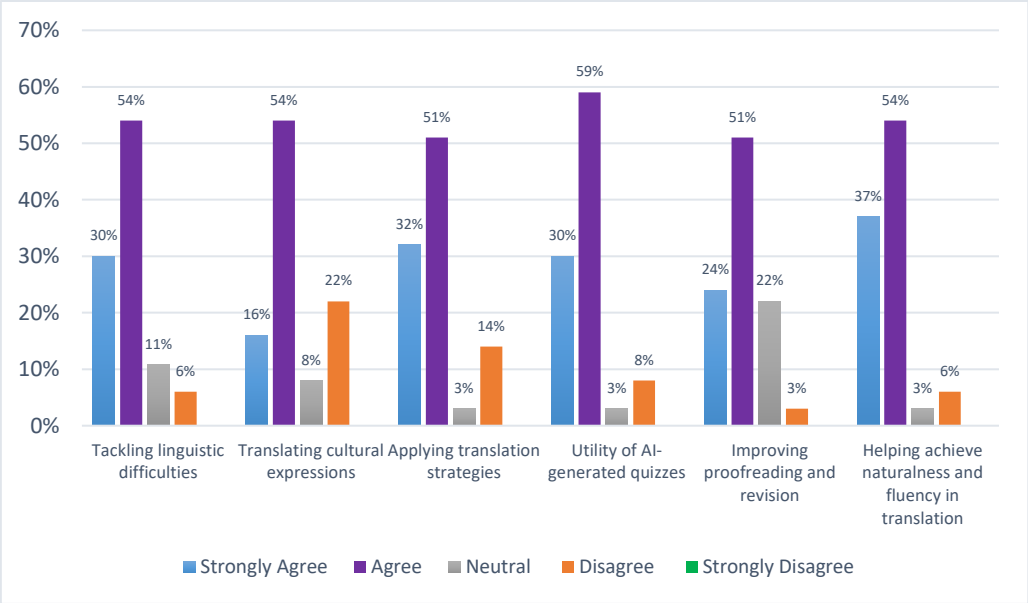
Table 12

Student Perceptions of AI Usefulness in Enhancing Translation Skills

No	Item	Mean	SD	% Agree/Strongly Agree
1	Tackling linguistic difficulties	4.08	0.80	83.8%
2	Translating cultural expressions	3.65	1.01	70.3%
3	Applying translation strategies	4.03	0.96	83.8%
4	Utility of AI-generated quizzes	4.11	0.81	89.2%
5	Improving proofreading and revision	3.97	0.76	75.7%
6	Helping achieve naturalness and fluency in translation	4.24	0.76	91.9%

Figure 4

Student Perceptions of AI Usefulness in Enhancing Translation Skills



As presented in Table 12 and Figure 4, participating students overwhelmingly perceived the AI-assisted training as highly effective for improving core translation competencies. The highest agreement was observed for AI usefulness in improving students' ability to produce natural and fluent translations (Item 6, 91.9% agreement, M = 4.24), closely followed by the utility of AI-generated quizzes for building translation proficiency (Item 4, 89.2% agreement, M = 4.11). High agreement was also recorded for AI tools' usefulness in tackling linguistic difficulties (Item 1, 83.8%, M = 4.08) and for applying translation strategies (Item 3, 83.8%, M = 4.03). Furthermore, a substantial majority (75.7% agreement, M=3.97) recognized AI benefits in improving proofreading and revision skills (Item 5). Finally, while still positive, students' perceptions of the AI enhancement of their ability to translate cultural expressions (Item 2, M = 3.65) received a comparatively lower mean score, though a strong majority (70.3%) still agreed or strongly agreed with the statement.

Table 13

Perceptions of AI General Pedagogical Benefits and Implications

No	Item	Mean	SD	% Agree/Strongly Agree
7	Encouraging peer collaboration	3.65	1.01	70 %
8	Creating interactive learning process	4.46	0.69	89.2%
9	Effectiveness of structured modules	4.30	0.70	91.9%
10	Utility of prepared prompts	4.27	0.65	89.2%
11	Importance of instructor guidance	4.30	0.81	89.1%
12	Fostering problem-solving independence	4.30	0.81	89.1%
13	Enhancing translation learning and teaching	4.35	0.82	83.8%

Figure 5

Perceptions of AI General Pedagogical Benefits and Implications

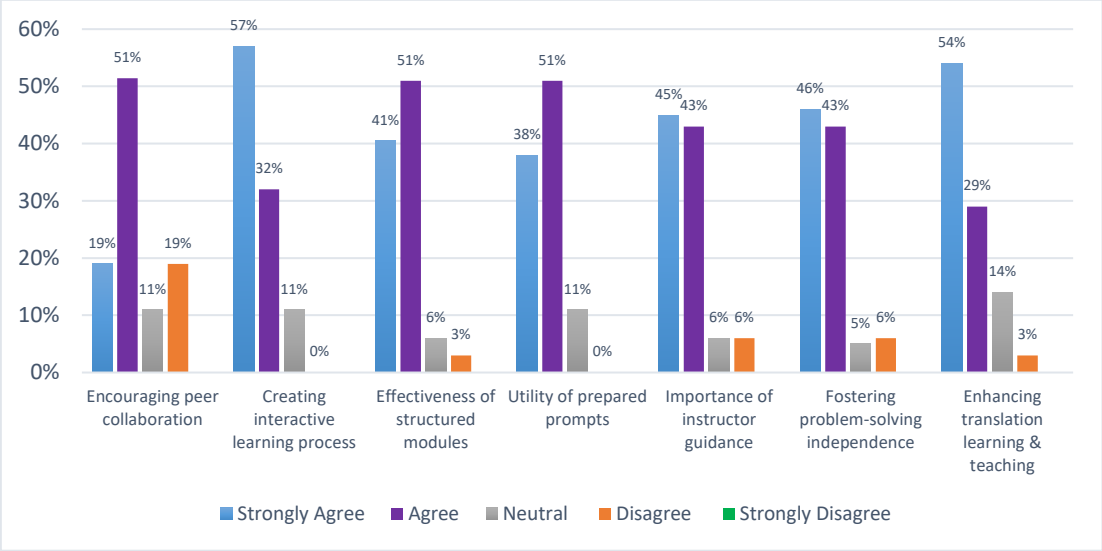


Table 13 and Figure 5 detail participants' perceptions of the broader pedagogical benefits of AI-assisted training. The highest mean rating was for AI role in creating an interactive learning process (Item 8, M=4.46), while the strongest consensus was on the effectiveness of structured training modules (Item 9, 91.9% agreement, M=4.30). This underscores the importance of pedagogical scaffolding, which was further emphasized by the high value placed on instructor guidance (Item 11, 89.1% agreement, M=4.30). Participants also strongly agreed on the utility of prepared prompts (Item 10, 89.2%, M=4.27) and reported that the intervention fostered learner autonomy in problem-solving (Item 12, 89.1%, M=4.30). A strong majority affirmed the overall potential of AI to enhance translation pedagogy (Item 13, 83.8%, M=4.35). In contrast, AI's role in encouraging peer collaboration received comparatively lower, though still positive, agreement (Item 7, 70%, M=3.65).

To sum, the perception findings confirm that students viewed the pedagogically oriented AI training as highly effective for translation skill development. Significantly, these findings underscore that the success of integrating AI (ChatGPT & Gemini) in translation education is reliant on a structured pedagogical planning, expert guidance, and well-designed prompts, rather than on the mere provision of the AI tools themselves. These findings will be further discussed and linked to the research questions in the discussion section.

6. DISCUSSION

This study explored the effectiveness of a pedagogy-driven AI training intervention on the English ⇄ Arabic translation skills of university translation majors. The following discussion interprets the quantitative and qualitative findings within the context of the existing literature, addressing the three research questions concerning overall proficiency, specific sub-skills, and pedagogical implications.

6.1.Generative AI Effectiveness in Enhancing Overall Translation Proficiency

In light of answering the research first question, the quantitative findings provide robust evidence for the effectiveness of pedagogically oriented AI training in improving overall translation skills. The experimental group, trained using guided AI tools (ChatGPT-4o & Gemini 2.5), demonstrated significantly superior post-test performance compared to the control group taught via traditional methods. This is evidenced by the substantial difference in post-test median scores (EG Md = 51 vs. CG Md = 37) and the significant results of the Mann-Whitney U test ($U = 105.000$, $p<.05$), which revealed a statistically significant difference between the two groups' overall translation skills post-test scores in favour of the experimental group. Crucially, both groups began at an equivalent baseline (Md = 30), strengthening the causal inference that the AI intervention was responsible for the differential learning gains. While the control group made modest progress (median gain of +7), the experimental group's marked improvement (median gain of +21) underscores the profound effectiveness of the AI-assisted pedagogy.

This significant difference is attributed to the pedagogical framework that strategically oriented AI potential toward specific learning objectives. By creating an interactive, engaging, and reflective learning context, the intervention facilitated a more productive learning experience. These findings align with previous research results indicating substantial improvements in learner's translation performance through AI-assisted training (Bakhov et al., 2024; Emara, 2024; Wang, 2024). Critically, the current study extends these previous studies. While previous studies often demonstrated AI potential in improving performance in limited tasks, the present research shows its pedagogical applications across broader and multiple complementary domains, including linguistic accuracy, cultural competence, and text-quality level revision. Practically, the findings reveal that a purposefully designed integration of generative AI with a sound pedagogical framework can unlock its multi-faceted roles in translation education.

By emphasizing AI process and product and engaging students in iterative learning circles, interacting with AI immediate feedback, and guided revision, the intervention promoted active knowledge construction through problem-solving and reflective learning practices. This approach explains the magnitude of the EG's learning gain, repositioning AI from a mere translation tool to a multipurpose pedagogical agent under instructor supervision. Thus, the study demonstrates that the transformative educational potential of generative AI in translation education is achievable through careful pedagogical planning and integration.

6.2.AI Effectiveness in Enhancing Specific Translation Sub-Skills

To address the second research question, a statistical analysis was conducted to determine the impact of the AI-assisted training on specific translation sub-skills. The results, measured by

the Wilcoxon Signed-Rank test, showed a statistically significant improvement ($p < .05$) across all three sub-skills for the experimental group after receiving the AI-assisted intervention.

Post-test median scores were higher for the experimental group than the control group in each domain: linguistic accuracy (EG Md = 18 vs. CG Md = 12), cultural competence (EG Md = 14 vs. CG Md = 10), and text quality revision (EG Md = 20 vs. CG Md = 15). Obviously, the gain in cultural competence was the smallest of the three, aligning with existing research (e.g., Khoshafah, 2023; Alqohfa & Sanad, 2025). This finding underscores that navigating cultural meaning, particularly in English-Arabic translation, remains a distinct challenge for Generative AI.

These variant outcomes clarify AI pedagogical role. The strong performance in linguistic accuracy and text revision highlights its effectiveness in fostering linguistic and quality-improvement skills. On the other hand, the modest gain in cultural competence reveals a key limitation of AI. Within the pedagogical framework, however, this very shortcoming was leveraged into a learning opportunity. By guiding students to analyse AI culturally variable outputs, the instructor facilitated a process that cultivates student's advanced cultural judgment.

Overall, the AI intervention's success demonstrates the necessity of systematic pedagogical integration. Using a constructivist framework that positioned AI as an interactive dialogue partner (Levin et al., 2025), the training modules under the instructor's supervision guided student's critical thinking and reflective learning. The iterative learning cycle, driven by students' interaction with immediate AI feedback, helped them foster understanding through active exploration and critical self-assessment. This process created an effective learning environment for skill development, highlighting the transformative role of pedagogically guided Generative AI.

6.3.Pedagogical Implications of Orienting Generative AI in Translation Education

In light of answering the research third question, the findings of this study highlight several practical pedagogical implications of integrating AI in translation education.

6.3.1. AI as a Translation Skill-Enhancement Tool

The significant post-test gains are reinforced by the experimental group's positive perceptions. The participants' high agreement (83.8%) about AI efficacy in helping them tackle linguistic difficulties corresponds with their improvement in linguistic accuracy scores (Md=18 post-test vs. 12 pre-test). This is also consistent with research findings by Yang et al. (2025) on ChatGPT's capacity for enhancing linguistic support. For cultural competence, student perceptions (70.3%) were notably lower than for other skills, revealing a critical issue in AI potential in this domain. Despite quantitative gains, cultural translation remains a challenging domain where AI utility is perceived as limited. This perception is consistent with established research on AI challenges with cultural nuances in English-Arabic translation (Khoshafah, 2023; Zaid & Bennoudi, 2023). On the other hand, these findings indicate that the developed pedagogical framework successfully enhances learning in an area of inherent AI weakness, underscoring the instructor's vital role in bridging this gap.

Furthermore, a strong agreement (89.2%) on the usefulness of AI-generated quizzes and immediate scoring confirms its efficacy as an automated formative assessor, a role validated in specialized domains like legal translation (Alghamdi & Alotaibi, 2025; Sadiq, 2025). High agreement percentages also affirm AI role as a collaborative assistant in revision, both for

enhancing the production of natural and fluent translations (91.9%) and improving proofreading skills (75.7%). This supports AI function within the study as both a 'proofreader & stylistic editor' and a 'collaborative assistant' for text quality revision and evaluation.

6.3.2. Broader Pedagogical Benefits

The findings revealed positive perceptions of broader pedagogical benefits of AI integration in translation education. The training was perceived as highly interactive (89.2% agreement), enhancing an engaging learning environment. The critical importance of pedagogical scaffolding was evident with 91.9% of students affirming the superior effectiveness of structured modules over independent AI use. This is consistent with research on instructor-mediated integration promoting engagement (Zhang & Doherty, 2025).

The crucial importance of pedagogically designed prompts was strongly supported with (89.2%) of students indicating that pre-designed prompts enhanced their translation skills, along with the role of AI in fostering learner autonomy as 89.1% of students reported increased autonomy in solving translation problems. Similarly, the indispensable role of the instructor was reaffirmed with (89.1%) of students agreeing that guided training with an instructor was more efficient than using AI on their own, underscoring the need for human oversight and supervision to provide contextualized and validated feedback. This is consistent with existing research findings (Xu et al., 2024; Sadiq, 2025; Alghamdi & Alotaibi, 2025). Finally, 83.8% of students confirmed the usefulness of guided AI tools (ChatGPT & Gemini) in enhancing translation learning and teaching, affirming the broader pedagogical potential of the guided AI in transforming translation education. These findings reveal the practical multi-faceted educational potential of generative AI in translation training when integrated into a solid pedagogical framework.

6.3.3. Optimizing AI-Structured Pedagogical Framework in Translation Education

A major significant contribution of this study in translation pedagogy is its empirically validated framework and its feasible application. This pedagogical framework redefines translation pedagogy by establishing an interactive, complementary partnership among student, AI, and instructor. It operationalizes constructivist principles (Li, 2014; Kynigos, 2015) by clearly defining distinct but complementary roles, as detailed in Table 14. According to this model, generative AI serves as a dynamic content generator and automated assessor; the student as an active knowledge constructor and reflective practitioner; and the instructor as the pedagogical designer and facilitator of metacognition. This partnership is enacted through defined interactive cycles, as illustrated in Figure 6, beginning with the instructor configuring the AI, which then engages the student in a formative feedback circle. The high student perception of this developed model (91.9% for structured modules, 89.1% for instructor's guidance) confirms its effectiveness in creating an effective and coherent learning ecosystem.

Therefore, the pedagogy-driven framework resolves a central dilemma in AI technology-enhanced learning by demonstrating that the instructor's role is not diminished but transformed into more crucial roles. This extends Vygotskian theory (Vygotsky, 2018), repositioning AI as a scaffold and the instructor as the essential guide for internalizing translation skills. Consequently, this model fosters a learning environment that simultaneously improves

translation proficiency and critical skill internalization, culminating in transformative outcomes for translation education.

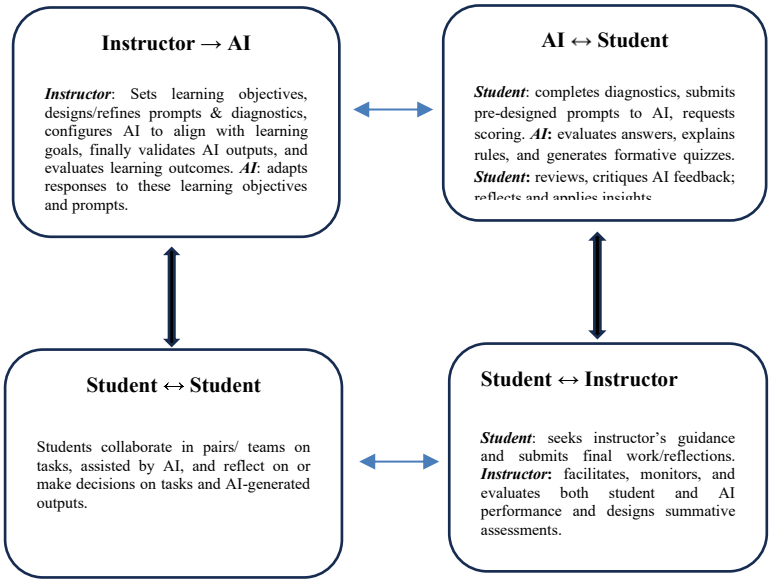
Table 14

Complementary Pedagogical Roles in an AI Driven Pedagogy Framework

Pedagogical Area	Generative AI's Role	Student's Role	Instructor's Role
Knowledge & Skill Delivery	Dynamic Content Generator	Active explorer, engager & knowledge constructor	Designer and director of Learning content
Feedback & Assessment	Automated Formative Assessor & Generator	Explorer &Reflective Practitioner	Metacognition Facilitator
Critical Thinking & Analysis	Subject of Evaluation	Critical Analyst	Critical Thinking Tutor
Practical Application & Creation	Collaborative Assistant	Learning Practitioner	Scenario Designer and Validator/ Final Judge of AI Output
Learning Environment & Scaffolding	Personalized Learning tutor	Self-Directed Learner	Scaffolding Provider & Moderator

Figure 6

AI Pedagogy Driven-Framework: Complementary Roles



6.3.4. Implications for Pedagogical Prompt Design and Translation Curriculum

The development of the AI training modules followed a rigorous, iterative cycle to ensure all outputs were aligned with specific learning objectives. This pedagogy-driven prompt engineering was the crucial success factor, transforming generative AI from a simple translation device into a structured pedagogical assistant. Targeted strategies such as few-shot prompting (Liang et al., 2023) and strategic persona assignment (He, 2024) were employed in line with learning objectives. For instance, using a persona prompt like 'Act as a translation tutor for an Arab student' contextualized interactions and tailored explanations. For tasks requiring conceptual precision, structured few-shot prompts with clear definitions and examples guided the AI to apply the correct analytical framework consistently.

This principled design generated student-friendly outputs, such as scored evaluations with rationales, facilitating a critical learning cycle where students could compare their answers, analyse feedback, and consider alternatives. Consequently, the pre-designed prompts were vital for a reflective pedagogy, directly supporting the development of analytical and self-regulated

translation skills. The high perceived usefulness of these prepared prompts (89.2%) reveals the importance of prompt engineering in AI pedagogy-driven framework, indicating a necessary prompt literacy.

Accordingly, this model provides a practical framework for educators, transforming generative AI into a consistent pedagogical partner capable of structuring a learning environment for effective translation skill development. These insights align with research emphasizing the critical role of well-designed prompt engineering in producing effective AI outputs for translation training (Alghamdi & Alotaibi, 2025; Yang et al., 2025).

6.3.5. Addressing Challenges in Pedagogy -Driven AI Integration

Although generative AI offers pedagogical benefits, its effective integration requires addressing key challenges. A major challenge is the lack of AI literacy on the part of students and instructors, specifically in designing pedagogical prompts. Translation programs should, therefore, hold mandatory training workshops at the start of the academic year. Instructors additionally need professional development in pedagogically grounded prompt engineering to maintain their central role in guiding AI-assisted tasks (Alghamdi & Alotaibi, 2025; Zhang & Doherty, 2025). Another challenge is the inherent variability of AI outputs and limitations, especially concerning cultural nuance (Xu et al., 2024; Sadiq, 2025). Mitigating these shortcomings requires combining robust prompt engineering training with critical reflection from instructors and students.

In general, this study has limitations that suggest fertile avenues for future research. The study sample, though robust for an experimental design (74 participants), was drawn from Yemen and Oman; replication across diverse Arab contexts would enhance generalizability. The multidimensional focus hindered a detailed comparative analysis of ChatGPT and Gemini performance across specific domains, leaving scope for future research to compare these two tools along with DeepSeek in specific translation areas, such as linguistic or cultural competence. Future studies should investigate AI role in developing stylistic variation and performance in specialized translation fields (e.g., literary, technical) in translation education context. Finally, longitudinal mixed-methods designs are recommended to examine the long-term pedagogical and cognitive impacts of such AI tools integration in translation education.

7. CONCLUSION

This study provides robust experimental evidence that a pedagogy-driven generative AI intervention (ChatGPT-4o & Gemini 2.5) significantly enhances English ⇄ Arabic translation skills. Employing a true experimental pre-test-post-test control group design (n=37 per group), it demonstrated that structured AI-assisted training was more effective than traditional instruction.

The study findings answer the primary research questions. The experimental group achieved significantly greater gains in overall translation proficiency (research first question), with consistent, statistically significant improvements across all three targeted sub-skills: linguistic accuracy, cultural competence, and text-level quality evaluation (research second question). This consistent progress led to the rejection of all null hypotheses. The quantitative results were strongly supported by participant perceptions, which affirmed the effectiveness of the guided AI framework for improving linguistic accuracy and quality evaluation

skills. However, perceived effectiveness was lower for cultural competence, a finding aligned with established literature on AI limitations with cultural nuances, which underscores the crucial role of instructor's supervision and oversight.

The effectiveness of this integration is supported by adhering to pedagogy-driven prompt engineering and a constructivist shift from AI product to process-oriented learning. The core theoretical contribution of this work is an empirically validated pedagogical model that enhances metacognition through a structured workflow and redefines the learning ecosystem by establishing specific, complementary roles. This creates a dynamic environment where students engage in a constructive 'give-and-take' with AI tools under the instructor's essential guidance, transforming the instructor's role from a translation corrector to a facilitative mentor and metacognitive tutor.

In answering the research third question, this study offers a practical, scalable framework for AI integration in translation pedagogy. It provides curriculum designers with an operational, evidence-based roadmap grounded in structured pedagogical design, instructor mediation, and student's reflective practice to transform generative AI into a multi-faceted pedagogical agent. Therefore, this research presents a critical rationale for moving beyond competition with AI toward designing sophisticated learning ecosystems where AI challenges are transformed into educational opportunities. This elevates the human instructor's roles of learning content design, mediation, supervision, fundamentally shifting AI from a potential threat into a driver for deeper, more effective learning in translation education and training.

RECOMMENDATIONS

Based on the study findings, the following strategic recommendations and initiatives are proposed:

Academic institutions should implement comprehensive annual workshops for instructors and students, dedicated to pedagogically-driven prompt engineering and the critical reviewing of AI outputs. These skills should be recognized as essential components of contemporary translation literacy and industry.

Translation curricula must be updated and revised to explicitly integrate AI interaction strategies and pedagogical potential into core learning objectives. Accordingly, instructor professional development should refocus on the design of AI-mediated learning experiences and the facilitation of critical reflection, firmly grounded in constructivist theory of learning.

To ensure broad effect, the study training modules could be updated and disseminated as a free, open-access online platform (e.g., hosted in collaboration with the Arab Observatory for Translation). This would serve as a living resource, enabling continuous improvement, regional adaptation, and the democratization of pedagogy-driven AI training across the Arab academic landscape.

Funding: This research received grant No. (2024/588) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing and Translation Commission in Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgements

The authors sincerely thank the participating students at the four universities for their valuable commitment as well as the academic leadership and staff for their cooperation in facilitating this study procedures. Thanks should go to the expert validators of the research instruments for their cooperation. We also thank and acknowledge the Arab Observatory for Translation and Translation Commission of Saudi Arabia for their funding support and their leading role in advancing translation research in the Arab world.

Ethical Approval

This study ethical approval was formally obtained from the concerned Deanships and Academic Departments at the 4 participating universities: Sana'a University, Yemeni Jordanian University, Al-Nasser University (Yemen) and University of Nizwa (Oman). Written informed consents were also obtained from all participating subjects in these universities before the study was conducted.

Data Availability Statement: The data presented in this research are available upon request from the corresponding author on reasonable request.

REFERENCES

Abu-Ayyash, E. A. (2017). Errors and non-errors in English-Arabic machine translation of gender-bound constructs in technical texts. *Procedia Computer Science*, *117*, 73–80. https://doi.org/10.1016/j.procs.2017.10.095

Al-Ali, L. R. (2025). The influence of AI on improving translation skills: A survey study. Wasit Journal for Human Sciences, 21(1 Pt1), 919–950. https://doi.org/10.31185/wjfh.Vol21.Iss1/Pt1.841

Aleedy, M., Atwell, E., & Meshoul, S. (2022). Towards deep learning-powered chatbot for translation learning. School of Computing, University of Leeds; College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University. https://eprints.whiterose.ac.uk/183815/

Alghamdi, F. A., & Alotaibi, H. (2025). Using AI in translation quality assessment: A case study of ChatGPT and legal translation texts. *Electronics*, 14(19), 3893. https://doi.org/10.3390/electronics14193893

Alharbi, W. (2024). Future translators' linguistic and non-linguistic competencies and skills in the age of neural machine translation and artificial intelligence: A content analysis. *International Journal of Linguistics, Literature and Translation*, 7(4), 124–143

Alqohfa, K. a. S., & Sanad, S. N. T. (2025). Linguistic and Cultural Challenges in Translating Idioms with Artificial Intelligence. *Journal of English Studies in Arabia Felix*, *4*(1), 34–49. https://doi.org/10.56540/jesaf.v4i1.116

Amaro, V., & João Pires, M. (2024). Found in translation, lost in education: Artificial intelligence's impacts on translation tertiary education in Macao. Asian Education and Development Studies. https://doi.org/10.1108/AEDS-01-2024-0012

Atlas, S. (2023). *ChatGPT for Higher Education and Professional Development: A guide to Conversational AI*. DigitalCommons@URI. https://digitalcommons.uri.edu/cba_facpubs/548

Bakhov, I., Bilous, N., Saiko, M., Isaienko, S., Hurinchuk, S., & Nozhovnik, O. (2024). Beyond the dictionary: Redefining translation education with artificial intelligence-assisted app design and training. *International Journal of Learning Teaching and Educational Research*, 23(4), 119–140. https://doi.org/10.26803/ijlter.23.4.7

Bhattacherjee, A. (2019). Social science Research: principles, methods and practices. In *University of Southern Queensland ePrints (University of Southern Queensland)*. https://doi.org/10.26192/q7w89

Brown, K. (2019). Personalized learning and artificial intelligence in the language classroom. *Journal of Educational Technology Research*, 28(3), 232–245

Campbell, D. T., & Stanley, J. C. (2015). *Experimental and Quasi-Experimental designs for research*. Ravenio Books.

Cheng, M. (2022). Practical exploration of English translation activity courses in colleges and universities under the background of artificial intelligence. *Advances in Education and Teaching International*, Special Issue. https://doi.org/10.1155/2022/4547342

Conover, W. J. (1999). *Practical Nonparametric Statistics* (3rd ed.). Wiley.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Ed-Dali, R. (2025). Assessing DeepSeek R1 and ChatGPT 4.5 in Arabic-English literary translation: performance, challenges, and implications. *Cogent Arts & Humanities*, 12(1). https://doi.org/10.1080/23311983.2025.2531183

Elen, J., Clarebout, G., Léonard, R., & Lowyck, J. (2007). Student-centred and teacher-centred learning environments: what students think. *Teaching in Higher Education*, *12*(1), 105–117. https://doi.org/10.1080/13562510601102339

Elkins, K. (2024). In search of a translator: using AI to evaluate what's lost in translation. *Frontiers in Computer Science*, *6*. https://doi.org/10.3389/fcomp.2024.1444021

Emara, E. a. E. M. (2024). Using AI Tools to Enhance Translation Skills among Basic Education English Major Students. *CDELT Occasional Papers in the Development of English Education*, *86*(1), 339–380. https://doi.org/10.21608/opde.2024.362829

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage.

Grassini, S. (2023). Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in *Educational Settings. Educ.* Sci. 2023, 13, 692. https://doi.org/10.3390/educsci13070692

Hajdu, Z., & Farkas, J. (2025). AI technology in technical translation courses. *Humán Innovációs Szemle*, *16*(1). https://doi.org/10.61177/hisz.2025.16.1.2

He, S. (2024). Prompting ChatGPT for translation: A Comparative analysis of translation brief and persona prompts. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2403.00127

Hellmich, E., & Vinall, K. (2021). FL instructor beliefs about machine translation: Ecological insights to guide research and practice. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *11*(4), 1–18. https://doi.org/10.4018/IJCALLT.2021100101

Kalpana, T. (2014). A constructivist perspective on teaching and learning: A conceptual framework. *International Research Journal of Social Sciences, 3*(1), 27–29. http://www.isca.me/IJSS/Archive/v3/i1/6.ISCA-IRJSS-2013-186.pdf

Kanglang, R. (2021). Student satisfaction with online learning during the pandemic: A review of empirical studies. *Journal of Education and Practice*, 12(2), 24–37.

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

Khasawneh, M. a. S., & Shawaqfeh, A. T. (2024). Breaking traditional boundaries in translation pedagogy; Evaluating how senior lecturers have incorporated digital tools to enhance translation teaching. *World Journal of English Language*, *14*(4), 154. https://doi.org/10.5430/wjel.v14n4p154

Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy. https://doi.org/10.21203/rs.3.rs-2814154/v1

Kiraly, D. (2014). *A social Constructivist approach to translator education: Empowerment from Theory to Practice*. Routledge.

Koka, N. A. (2024). The integration and utilization of artificial intelligence (AI) in supporting older/senior lecturers to adapt to the changing landscape in translation pedagogy. *Migration Letters*, 21(S1), 59-71. https://www.migrationletters.com

Kornacki, M., & Pietrzak, P. (2024). *Hybrid workflows in translation: Integrating GenAI into translator training* [Print]. Routledge.

Kynigos, C. (2015). Constructionism: Theory of Learning or Theory of Design? In *Selected Regular Lectures from the 12th International Congress on Mathematical Mathematical Education* (pp. 417–438). Springer. https://doi.org/10.1007/978-3-319-17187-6_24

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: a systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, *22*(1). https://doi.org/10.1186/s41239-025-00503-7

Levin, I., Semenov, A. L., & Gorsky, M. (2025). Smart Learning in the 21st Century: Advancing constructionism across three digital epochs. *Education Sciences*, *15*(1), 45. https://doi.org/10.3390/educsci15010045

Li, J. (2014b). A Tentative Approach of Translation Education under the Constructivism Theory. *Advances in Social Science, Education and Humanities Research/Advances in*

Social Science, Education and Humanities Research. https://doi.org/10.2991/icssr-14.2014.196

Liang, Y., Wang, J., Zhu, H., Wang, L., Qian, W., & Lan, Y. (2023). Prompting Large Language Models with Chain-of-Thought for Few-Shot Knowledge Base Question Generation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2310.08395

Mohammed, S. Y., & Aljanabi, M. (2024b). Advancing Translation Quality Assessment: Integrating AI models for real-time feedback. *EDRAAK*, *2024*, 1–7. https://doi.org/10.70470/edraak/2024/001

Mohsen, M. A. (2024). Artificial intelligence in academic translation: A comparative study of large language models and Google Translate. *Psycholinguistics*, *35*(2), 134–156. https://doi.org/10.31470/2309-1797-2024-35-2-134-156

Newmark, P. (1988). *A textbook of translation*. Prentice-Hall International.

Omar, L. I., & Salih, A. A. (2024). Systematic Review of English/Arabic Machine Translation Postediting: Implications for AI Application in Translation Research and Pedagogy. *Informatics*, *11*(2), 23. https://doi.org/10.3390/informatics11020023

Pallant, J. (2020). *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS* (7th ed.). McGraw-Hill.

Richards, J. C., & Schmidt, R. (2001). *Longman Dictionary of Language Teaching and Applied Linguistics* (3rd ed.).

Sacristán, A. I., Baafi, R. a. K., Kaminskienė, L., & Sabin, M. (2018). *Constructionism in upper secondary and tertiary levels*. https://hdl.handle.net/20.500.12259/59905

Sadiq, S. (2025). Evaluating English-Arabic translation: Human translators vs. Google Translate and ChatGPT. *Journal of Languages and Translation*, *12*(1), 67–95. https://doi.org/10.21608/jltmin.2025.423147

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Cengage Learning.

Somers, H. (2021). Neural machine translation in language learning and teaching: potentials and pitfalls. *Lang. LearnJ ,*, *12*(2), 188–208.

Thowaini, B., & Qassem, M. (2024). The impact of an industry-driven translation training program on the performance of trainee translators. *Cogent Arts & Humanities*, 11(1), 2339618. https://doi.org/10.1080/23311983.2024.2339618

Tian, X. (2024). Personalized translator training in the era of digital intelligence: Opportunities, challenges, and prospects. *Heliyon*, (10), e39354. https://doi.org/10.1016/j.heliyon.2024.e39354

Vygotsky, L. S. (2018). *Mind in society: Development of Higher Psychological Processes*. Harvard University Press.

Wang, Y. (2023). Artificial intelligence technologies in college English translation teaching. *Journal of Psycholinguistic Research, 52*(5), 1525–1544. https://doi.org/10.1007/s10936-023-09960-5

Wang, Y. (2024). Preparing translators for the age of generative AI: Revising educational curricula to prioritize essential skills in an AI-dominated landscape. *Advances in Humanities Research, 9*. EWA Publishing. Published online: October 21, 2024. DOI:10.54254/2753-7080/9/2024109

Xu, S., Sun, Y., Su, Y., & Liu, K. (2025). Translation Pedagogy in the AI Era. In *Translation Studies in the Age of Artificial Intelligence* (pp. 100–124). Routledge. https://doi.org/10.4324/9781003482369-6

Xu, Y. (2024). AI empowered undergraduate translation teaching and professional improvement of translation teachers. *International Journal of Education and Humanities*, *17*(2), 253–256. https://doi.org/10.54097/cytrjn11

Yang, C., Chen, J., Hou, S., Wider, W., & Yi, W. L. (2025). How AI benefits student translators: An exploratory study on the impact of ChatGPT feedback on translation proficiency. *Forum for Linguistic Studies*, *7*(10). https://doi.org/10.30564/fls.v7i10.11080

Zaid, A., & Bennoudi, H. (2023). AI vs. human translators: Navigating the complex world of religious texts and cultural sensitivity. *International Journal of Linguistics, Literature and Translation, 6*(11), 173–182. https://doi.org/10.32996/ijllt.2023.6.11.21

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00316-7

Zhang, J., & Doherty, S. (2025). Investigating novice translation students' AI literacy in translation education. *The Interpreter and Translator Trainer*, *19*(3–4), 234–253. https://doi.org/10.1080/1750399x.2025.2541478

About the Authors

Redhwan Alsharef *is an Assistant Professor of English Language and Comparative Literature in the Department of English Language at Hajjah University, Yemen. As a translation practitioner, translation instructor and expert in translation pedagogy, he has extensive experience teaching translation courses and developing translation programs across several universities. His research and professional interests include developing technology-enhanced, pedagogy-driven translation applications and programs, exemplified by his Translation Companion App (available on Google Play: link).*

ORCID: 0009-0004-0584-287X

Faisal Aldawli *is an Assistant Professor of Translation Studies working as the head of translation Unit in the Community Center for Translation and Language Teaching at Sana'a University. He has been teaching translation courses in undergraduate programs in Sana'a University and private universities for over a decade.*

***ORCID**: 0009-0001-1151-597X*

Khaled Alahdal *is an Associate Professor of English Language and Literature in the Department of Foreign Languages (DFL) in the University of Nizwa (UoN), Sultanate of Oman. He has over 15 years of experience in teaching literature and translation courses.*

Nuha Shamsan *has a Master's Degree in Translation. She is a researcher, university lecturer and freelance translator. She has an extensive experience of translating numerous documents across various fields.*