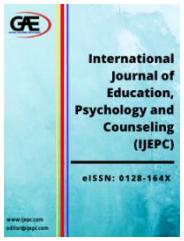
International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139



INTERNATIONAL JOURNAL OF EDUCATION, PSYCHOLOGY AND COUNSELLING (IJEPC)
https://gaexcellence.com/ijepc

INTEGRATING AI AND HUMAN TRANSLATION IN LITERARY TRANSLATION PEDAGOGY: A STRUCTURED PEDAGOGICAL SEQUENCE FOR CRITICAL REFLECTIVE JUDGEMENT

Sulhah Ramli[1]*, Hishomudin Ahmad[2], Muhammad Izuan Abd Gani[3] , Aishah Isahak[4]

[1]Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia
sulhah@usim.edu.my https://orcid.org/0000-0002-1881-5617
[2] Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia
hishomudin@usim.edu.my https://orcid.org/0000-0002-0435-7897
[3] Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia
izuanis@usim.edu.my https://orcid.org/0009-0003-6743-7266
[4] Fakulti Pengajian Bahasa Utama, Universiti Sains Islam Malaysia, Nilai, Negeri Sembilan, Malaysia
aishah_isahak@usim.edu.my https://orcid.org/0000-0003-0706-9337
*Corresponding Author

Article Info:

Article history:
Received date: 23.07.2026
Revised date: 17.08.2026
Accepted date: 06.09.2026
Published date: 20.09.2026

To cite this document:

Ramli, S., Ahmad, H., Gani, M. I. A., & Isahak, A. (2026). Integrating AI and Human Translation In Literary Translation Pedagogy: A Structured Pedagogical Sequence For Critical Reflective Judgement. *International Journal of Education, Psychology and Counselling, 11*(64), 1127-1139

Abstract:

Generative artificial intelligence (GenAI) gives student translators additional renderings to consider, each of which still requires human evaluation of meaning, style, and cultural context. This article examines how critical reflective judgement becomes visible in students' decisions during GenAI-assisted literary translation. The dataset comprises six anonymised student portfolios representing 23 distinct passages and 40 student–passage cases from the *Muʿallaqah of Zuhayr*, translated into Malay or English. Students produced an initial literal translation, consulted ChatGPT and Gemini using lecturer-provided prompts, prepared a refined version, and justified their decisions in writing. Comparative analysis of successive versions and reflective commentaries traced how students addressed semantic, stylistic, and cultural problems and, where AI-generated alternatives were identifiable, whether they accepted, modified, or rejected them. Critical reflective judgement was clearest when students tested AI suggestions against contextual or literary evidence, assessed their adequacy, and explained why particular renderings were retained, revised, or rejected. The portfolios make the intervening processes of comparison, evaluation, and decision-making more explicit within the staged translation task. They inform a seven-stage pedagogical sequence of translation, AI consultation, comparison, evaluation, decision-making,

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

refinement, and justification. Based on this bounded dataset, the sequence is an empirically informed pedagogical proposition, not a validated model of instructional effectiveness.

DOI: 10.35631/IJEPC.1164059

Keyword:

Generative AI; Literary Translation Pedagogy; Reflective Judgement; Translator Agency; Arabic Literary Translation

Introduction

Generative artificial intelligence (GenAI) is reshaping how translation is learned, revised, and evaluated. Translation technologies have long formed part of human-computer interaction, and research has cautioned against treating them as substitutes for translator judgement (O'Brien, 2012). Work on literary machine translation has also highlighted ethical concerns and the continuing importance of the translator's voice (Kenny & Winters, 2020). Recent studies report that students value GenAI for efficiency, idea generation, and linguistic support while remaining concerned about accuracy, over-reliance, accountability, and the need for pedagogical guidance (Zhang et al., 2025). The educational value of AI feedback depends on how actively learners process and evaluates it (Xu et al., 2025), especially because students' perceptions of AI translation tools do not always entail sufficiently critical assessment (Romaniuk-Cholewska, 2025). Human-centred approaches treat AI as a means of augmenting learners' interpretive agency and decision-making without displacing them (Jiménez-Crespo, 2025).

Literary translation makes these concerns particularly consequential. Successful performance depends on interpretive adequacy as well as fluency: translators must negotiate meaning, imagery, tone, style, and cultural implication and decide which effects to retain, reconfigure, or make explicit for a new readership (Boase-Beier, 2020; Kenny & Winters, 2020). These choices draw on strategic translation competence, including the ability to identify problems, activate relevant knowledge, compare alternatives, and select procedures appropriate to the communicative purpose (PACTE Group et al., 2018; Angelone, 2010). Classical Arabic poetry intensifies the challenge because compact expressions may presuppose historical practices, tribal relations, material culture, geographical knowledge, and culturally situated imagery. Contextual and environmental relations (Risku, 2002) shape translation, while Arabic poetic

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

meaning frequently requires attention to cultural and nonverbal dimensions that cannot be recovered through word-level correspondence alone (Lahiani, 2025).

Current scholarship shows the need for critical, human-centred engagement with AI, yet offers less clarity on how such judgement can be made visible and teachable within a staged literary-translation process. Revision alone is not evidence of reflection. A student may adopt a fluent AI alternative without showing the underlying translation problem or considering what the change gains and loses. Portfolios, written commentaries, and structured self-reflection can reveal problem-solving processes and support metacognitive regulation in translator education (Massey & Ehrensberger-Dow, 2011; Pietrzak, 2019; Eskelinen & Pakkala-Weckström, 2016). Further process-oriented evidence is needed on how students working with culturally dense literary texts compare their initial renderings with GenAI alternatives and justify decisions to accept, modify, or reject them.

Six student portfolios from BAL4523 Literary Translation course at Universiti Sains Islam Malaysia, anonymised as P1–P6, form the dataset. They document translations of selected passages from the *Muʿallaqah* of *Zuhayr* into Malay or English through an initial literal translation, consultation with ChatGPT and Gemini using lecturer-provided prompts, an interpretative and refined translation, and written explanation or justification. Critical reflective judgement is operationalised through observable decisions across these staged products: recognising semantic, stylistic, or cultural problems; comparing alternatives; identifying gains and losses; deciding whether to accept, modify, or reject a proposed solution; refining the target text; and linking the final wording to a reasoned justification.

The inquiry traces critical reflective judgement across successive translation stages and uses the observed decision patterns to inform a pedagogical sequence for accountable human–AI literary translation. It addresses three questions; RQ1: How do student translators handle semantic, stylistic, and cultural difficulties across successive translation versions? RQ2: How do they explain and justify their choices in relation to meaning, literary style, and cultural context? RQ3: How do their decisions change after consulting ChatGPT and Gemini, particularly when they accept, modify, or reject identifiable AI-generated alternatives? Attention centres on the reasoning behind translation choices, not on final-text fluency alone. Empirically, the portfolios show how students formulate and defend decisions in AI-assisted translation of classical Arabic literature. Pedagogically, they make comparison, evaluation, and decision-making more explicit within the task's existing stages of translation, AI consultation, refinement, and justification.

Literature Review

Literary Translation as Interpretive and Cultural Mediation

Literary translation involves a sequence of problem-solving decisions in which formal similarity alone may not produce an adequate target text. Translation competence entails recognising when an available equivalent is insufficient, identifying the nature of the problem, activating relevant knowledge, and selecting a solution that serves the communicative purpose of the target text (PACTE Group et al., 2018). In literary contexts, the problem may involve referential meaning, imagery, tone, social implication, or the relationship between an expression and the cultural world presupposed by the source text. These demands make stylistic interpretation central to literary translation (Boase-Beier, 2020).

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

In Arabic poetry, material objects, landscapes, tribal practices, and conventional images may carry meanings that a dictionary gloss alone cannot recover. Lahiani (2025) shows how cultural and nonverbal dimensions shape the translation of Arabic poetry through contextually situated signs and associations. Risku's (2002) account of translation as an activity embedded in contextual and environmental relations offers a compatible view of situated meaning. Students therefore need opportunities to make the grounds of their interpretations explicit, rather than being assessed solely on the apparent fluency of a final target text.

Translation Problem Solving and Reflective Judgement

Reflective judgement in translation is closely connected with the management of uncertainty. Schön's (1983) account of reflection-in-action provides a foundational view of professional learning in which practitioners interrogate problems while acting instead of merely applying fixed rules. Angelone (2010) describes translation problem solving as a metacognitive process in which translators recognise uncertainty, select strategies for resolving it, and monitor the adequacy of emerging solutions. Translator-training research shows that structured self-reflection can scaffold metacognitive regulation and make learning processes more explicit (Pietrzak, 2019). Written commentary and other forms of process evidence also reveal how students understand and justify their choices (Massey & Ehrensberger-Dow, 2011).

Strategic problem solving occupies a central place in translation performance (PACTE Group et al., 2018). Competence-based translator education uses purposeful tasks and assessment to develop and observe translation competence (Hurtado Albir, 2015), while social-constructivist approaches treat learners as active participants who build translation knowledge through authentic, collaborative, and reflective activity (Kiraly, 2000). Reflective commentaries, portfolios, and constructively aligned assessment can also trace the development of professional competences (Eskelinen & Pakkala-Weckström, 2016). A revised version becomes educationally informative when the decision behind the change can be followed.

Generative AI and Translation Revision

GenAI introduces another source of translation alternatives. Student engagement with AI-driven feedback during revision is uneven, and its educational value depends on how learners process, evaluate, and use it (Xu et al., 2025). Research on translation as human–computer interaction similarly emphasises translators' active engagement with technological tools and cautions against substituting technology for human judgement (O'Brien, 2012). The pertinent pedagogical question is how a student responds to an AI candidate: by adopting, modifying, challenging, or using it to prompt further investigation.

Students' critical evaluation of AI translation tools is central to Romaniuk-Cholewska's (2025) account. In literary translation, fluent generative output may conceal interpretive simplification, cultural under-specification, or stylistic overstatement. Machine translation can also affect the visibility and expression of the literary translator's voice, which makes explicit human judgement especially important in literary workflows (Kenny & Winters, 2020). An assignment that requires only an AI translation treats the technology as an answer generator. Comparison and justification instead make the generated alternative an object of inquiry.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

AI Alternatives, Epistemic Vigilance, and Translator Agency

Epistemic vigilance provides a sensitising lens for considering why generated alternatives require evaluation beyond surface plausibility. Sperber et al. (2010) use the term for processes through which communicated information is assessed for reliability and believability. In the present analysis, it does not imply that an internal cognitive state can be inferred from portfolio data. The concept instead identifies a modest educational requirement: students should examine the evidential and contextual basis of a generated translation before allowing it to shape the final target text.

Translator agency becomes visible when learners take responsibility for evaluating available alternatives. The number of changes made to AI output is not, by itself, a measure of agency. Human-computer interaction research in translation stresses the need to design technological workflows around translators instead of reducing them to passive operators (O'Brien, 2012). Critiques of digitally managed translation work likewise warn against systems that diminish professional autonomy and meaningful decision-making (Moorkens, 2020). Here, agency is evident when a student identifies a problem, compares alternatives, decides what to retain or alter, and justifies the final wording in relation to the source text and its cultural context. A reasoned, contextually supported acceptance of an AI suggestion can be as agentive as its rejection.

Positioning the Present Study

Research on AI feedback and translation revision points to the need for critical evaluation of AI translation tools (Xu et al., 2025; Romaniuk-Cholewska, 2025). The present focus is a staged literary-translation process in which students translate first, consult AI, compare and refine their work, and document the reasoning behind their decisions.

Successive versions and commentaries show how students undertake semantic, stylistic, and cultural mediation and account for changes made after AI consultation. This process-oriented account makes comparison, evaluation, and decision-making more explicit within a structured sequence for accountable human-AI interaction in literary translation. Its focus on classroom activity and learner decision-making accords with task-based approaches to translator education (González Davies, 2004). The study neither tests the sequence experimentally nor claims that one AI system is superior to another.

Methodology

A qualitative descriptive design was used to examine successive translation versions alongside students' written reflections. The dataset comprised six undergraduate portfolios from a Literary Translation course at Universiti Sains Islam Malaysia, with participants anonymised as P1-P6. The portfolios covered 23 distinct passages from the *Muʿallaqah* of *Zuhayr* and generated 40 student-passage cases. Students translated the assigned texts into Malay or English according to the task requirements.

All six portfolios available for the analysed activity were included because they contained the staged translation products and reflective commentaries needed to trace changes and their accompanying rationales. The corpus is bounded to this instructional setting and was not assembled for statistical generalisation. It supports the identification of decision-making

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139
patterns within these portfolios, not estimates of their prevalence among translation students more generally.

Translation and AI-Consultation Procedure

AI consultation followed an initial human attempt. Students first produced a literal or closely source-oriented translation, then consulted ChatGPT and Gemini using task instructions and prompts provided by the lecturer. Consultation took place before they prepared the interpretative version. Students subsequently developed interpretative and refined or contextual versions and explained difficult expressions, strategy choices, cultural considerations, and final translation decisions.

This ordering permits comparison between each student's pre-consultation rendering and later decisions after exposure to AI alternatives and further contextual work. The later versions are not assumed to have been caused by AI consultation. Raw AI outputs and interaction histories were not preserved consistently across the portfolios, so direct claims about AI influence are restricted to cases in which the consulted alternative is identifiable in the submitted material.

Unit of Analysis

The primary analytical unit was the translation-decision episode. Each episode linked a source expression or passage segment to the student's successive target renderings and, where available, the corresponding reflective explanation and identifiable AI alternative. It supported consideration of the problem the student appeared to recognise, changes across versions, their semantic, stylistic, or cultural consequences, and the justification for the final choice.

Decision episodes were examined across all 40 student-passage cases. Particular attention was given to culturally embedded material involving tribal compensation and collective responsibility, geographical reference and social protection, camel and material culture, and culturally situated poetic imagery. These domains organised recurring interpretive demands but were not treated as exhaustive categories of the *Muʿallaqah* of *Zuhayr*.

Analytical Procedure

The analysis proceeded in three linked passes. First, successive translations were aligned at the relevant expression or passage level. Changes were described by their effects on meaning and representation, including lexical expansion, contextual specification, retention with explanation, stylistic reshaping, and cultural reinterpretation. Analytical attention remained on the consequence of each change, not on strategy labels alone.

Second, reflective comments were read alongside the corresponding textual changes for evidence of problem recognition, uncertainty, contextual interpretation, evaluation of alternatives, reader consideration, and explicit justification. A statement that an expression had no direct equivalent was treated as analytically meaningful only when the accompanying explanation specified what the available target term failed to convey or why further mediation was needed.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

Third, when an AI-generated candidate was identifiable, its relation to the student's later translation was described as acceptance, modification, rejection, or indeterminate. The indeterminate category applied when the portfolio did not preserve enough evidence to establish that relation. No frequency comparison between ChatGPT and Gemini was conducted, and revisions were not attributed to either system without documentary support.

Cross-Case Analysis and Ethical Treatment

Within-case observations were compared across P1-P6 to identify recurring decision patterns and contrasting cases. Cross-case comparison showed whether an interpretive issue appeared in more than one target language or participant portfolio and helped distinguish recurrent pedagogical patterns from peculiar solutions. Interpretive claims remained close to the submitted source-target evidence and, where available, the participant's own explanation.

Verbal consent for the research use of the portfolio materials was obtained from all six participants. Cases were anonymised as P1-P6, and no personally identifying student information is reported. The evidence consists of submitted educational artefacts and their documented translation decisions.

Findings

Semantic, Stylistic, and Cultural Mediation Across Stages

Revisions across the six portfolios were selective. Students retained close wording when it remained adequate and intervened when an initial rendering obscured a culturally relevant relation, weakened a poetic effect, or activated a misleading modern association. The successive versions reflect targeted mediation rather than a general movement away from literal translation or a uniform preference for expansion.

P2's treatment of the *diyah*-related expression in *Bait* 43 illustrates this pattern. The early rendering stayed close to the literal action of 'tying'. A later version moved towards compensation, then to a culturally marked rendering that retained *diyat* and restored the relevance of camels within the compensation practice. P2 described the concept as having '*tiada padanan langsung dalam BM*' and distinguished the surface physical action from the social practice of formally delivering camels as blood compensation. The key analytical development was the recognition that a general Malay equivalent did not recover the institutional relation encoded in the source; the added information was not an end in itself.

Comparable mediation appeared across other portfolios. P1 and P3 interpreted references to inherited wealth and marked young camels in relation to ownership, lineage, status, and social value, extending beyond animal description alone. P5 and P6 treated place names and protection-related expressions as requiring contextual knowledge of route, covenant, risk, and social order. In the treatment of *al-ʿihn* and *ḥabb al-Fanā*, visual description was refined through attention to coloured textile imagery, camel-litter ornamentation, and the cultural scene presupposed by the poem. Cultural mediation in these cases made a relation recoverable without replacing a source-culture item with a fully domestic substitute.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

Stylistic intervention was similarly selective. Students reorganised wording, intensified or reduced imagery, and altered target-language cadence when an initial version conveyed propositional content but did not function adequately as literary language. These changes were most defensible when the reflective note identified the intended effect and the revised wording remained anchored in the source. Some attractive elaborations lacked sufficient support, showing that literary fluency can also create a risk of over-interpretation.

Reflective Justification and Translator Agency

The reflective commentaries varied substantially in explanatory depth. Basic comments named a difficulty or translation procedure. Stronger comments identified what an available rendering failed to convey, considered the consequence for the reader, and linked the selected procedure to a specific textual problem. This difference underpins the operationalisation of critical reflective judgement: a strategy label such as functional equivalent or modulation does not by itself explain why a decision was appropriate.

Several portfolios demonstrate this deeper form of justification. When students encountered culturally loaded legal or social expressions such as *gharāmah*, collective responsibility, or protection, they moved from broad modern terms toward explanations that recovered the relevant social function. In cases involving place names or source-culture institutions, students sometimes retained the Arabic term or proper name and used contextual explanation to preserve specificity. In other cases, a descriptive or functional rendering was preferred because retention would have added opacity without interpretive benefit.

The clearest evidence of translator agency appeared when a written rationale corresponded directly to an observable textual change. A student who identified a missing cultural component and restored it in the refined version created a traceable link between diagnosis and intervention. By contrast, a commentary that merely asserted that a choice was 'better' or named a strategy without a visible textual consequence offered weaker evidence of critical reflective judgement. In this dataset, agency is best understood as accountable choice, not independence from all external assistance.

Student Decisions After AI Consultation

AI consultation expanded the set of available alternatives, but the portfolios did not show a simple pattern of acceptance or rejection. In identifiable episodes, students sometimes retained an AI lexical or syntactic proposal after comparison supported its adequacy. Elsewhere, they modified a generated rendering by restoring cultural specificity, adjusting literary tone, or adding contextual information. They rejected alternatives that flattened a culturally significant distinction, activated an unsuitable contemporary meaning, or introduced stylistic force without sufficient support in the source.

The most informative episodes treated the AI candidate as an object of comparison, not as a final answer. Students who asked what the generated version omitted, what cultural assumption it made, or why it sounded persuasive provided fuller accounts of subsequent revisions. Where the portfolio did not require students to articulate the basis of a decision, a fluent alternative could be incorporated with little evidence of critical evaluation.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

The documentary limits of the portfolios constrains these observations. Complete raw outputs from ChatGPT and Gemini were not consistently retained, so the record does not permit quantification of accepted, modified, or rejected suggestions or comparison of the two systems' performance. Some later revisions remain indeterminate with respect to direct AI influence. The findings therefore concern post-consultation evaluation within the pedagogical process, not system-level accuracy or causal effects of GenAI use.

Discussion

Reflective Judgement as Translation Problem Management

The portfolios present reflective literary translation as a form of structured problem management. The strongest revisions began with recognition that an apparently fluent or literal rendering failed to recover a relevant semantic, stylistic, or cultural relation. This pattern resembles Angelone's (2010) account of uncertainty management, in which translators identify uncertainty, seek a means of resolving it, and monitor the resulting solution. Written reflection made parts of that regulation visible because learners had to state what was problematic and why they preferred the final wording.

PACTE Group et al's (2018) strategic dimension of translation competence helps explain the same pattern. A change in wording alone did not demonstrate competence. Stronger strategic performance connected the change to a diagnosed problem and the purpose of the target text. By preserving successive versions and their rationales, the staged portfolio provides evidence for assessing that relationship.

AI Alternatives and Epistemic Vigilance

Post-consultation patterns accord with research showing that the educational value of AI feedback depends on students' active engagement during revision (Xu et al., 2025). Translation students also report valuing GenAI for generating alternatives and supporting practice while recognising risks involving accuracy, dependency, transparency, and accountability (Zhang et al., 2025). For the present analysis, the pedagogically relevant distinction is between evaluated and unevaluated wording, whether human or machine generated. An AI alternative may create productive contrast, prompt a search for contextual evidence, or draw attention to a problem that was not apparent in the first draft.

The portfolio evidence also supports calls for explicit critical routines when students evaluate AI translation tools (Romaniuk-Cholewska, 2025). GenAI fluency can make an under-specified or culturally flattened translation appear complete. Human-centred AI scholarship in translation education advocates models that preserve human agency, critical judgement, and meaningful interaction with technology without training students to imitate automated behaviour (Jiménez-Crespo, 2025). Epistemic vigilance supplies a useful sensitising principle: communicated information should be assessed before it is relied upon (Sperber et al., 2010). Students therefore need to ask what supports an AI alternative, what it presupposes, and what follows if they adopt it.

International Journal of Education, Psychology and Counseling
IJEPC
EISSN : 0128-164X
Volume 11 Issue 64 (September 2026) PP. 1127-1139

Difficulties involving tribal institutions, material culture, geographical reference, and literary imagery recurred across the portfolios, illustrating the demands of classical Arabic poetry in AI-assisted translation pedagogy. Lahiani (2025) emphasises that translating Arabic poetry requires attention to cultural and nonverbal dimensions beyond lexical correspondence. The cases show how an apparently small lexical choice can reorganise the cultural scene presented to the target reader.

Proportionate mediation does not require every culture-bound item to receive maximal explanation. Students must determine which component is necessary for interpretation, whether the target wording recovers it, and whether a gloss, retained term, descriptive equivalent, or stylistic reconfiguration offers the least intrusive solution. Critical reflection checks for both under-translation and speculative enrichment.

An Empirically Informed Pedagogical Sequence

The observed patterns inform a seven-stage sequence for organising AI-assisted literary translation: translation, AI consultation, comparison, evaluation, decision-making, refinement, and justification. Translation preserves the initial human attempt before exposure to generated alternatives. AI consultation introduces ChatGPT, Gemini, or another permitted AI system as a source of candidate renderings. Comparison places the student's version beside the generated alternatives, and evaluation considers their effects on semantic meaning, literary style, cultural context, and reader interpretation. Decision-making requires an explicit choice to accept, modify, or reject an available solution. Refinement produces the target text after that judgement, while justification connects the final wording to the identified problem and the evidence used to resolve it. Translation, AI consultation, refinement, and written justification were already built into the task; the portfolio analysis makes comparison, evaluation, and decision-making within that workflow more explicit.

The sequence keeps a human evaluator in the process without assuming that human judgement is automatically correct. AI supplies additional variation for comparison, while the learner remains responsible for contextual research and the final translation decision. The sequence is an empirically informed pedagogical proposition derived from the observed portfolio process, not a validated model. Its effectiveness as an instructional intervention requires separate testing.

Scope and Limitations

The evidence is bounded by six portfolios from one undergraduate course and 23 distinct passages. Reflective commentary is retrospective and cannot reproduce every cognitive step taken during translation. Incomplete preservation of raw AI interactions also prevents exhaustive tracing of system-specific influence and rules out a valid comparative assessment of ChatGPT and Gemini.

These limits set the boundaries of the process evidence. The pedagogical sequence rests on observable patterns of mediation and justification within the six portfolios. Future research could test it prospectively by requiring systematic preservation of AI prompts and outputs, applying the procedure across larger cohorts and literary genres, and examining whether

structured comparison and justification improve translation quality or critical reflective judgement over time.

Conclusion

Students' revisions show that critical reflective judgement was most evident when they recognised a semantic, stylistic, or culturally situated problem and could explain how their chosen rendering addressed it. This pattern appeared across six portfolios comprising 40 student-passage cases from 23 distinct passages of the *Muʿallaqah* of *Zuhayr*. Some reflections were brief. Others were more developed, particularly when students compared alternatives, explained the consequence of a choice, and linked that reasoning to the final wording. Critical reflective judgement was therefore located in the evaluation and justification of particular decisions, rather than in the number of revisions made or in resistance to AI suggestions.

AI consultation widened the range of alternatives available for consideration. It did not remove the need for judgement. Fluent output still had to be assessed against semantic, stylistic, and cultural demands, and the incomplete preservation of AI outputs limits what can be claimed about system comparison, acceptance or rejection rates, and the causes of later revisions. What can be traced more securely is how students assessed identifiable AI-generated alternatives while retaining responsibility for the final translation choice.

The recurring decision processes support a seven-stage pedagogical sequence: translation, AI consultation, comparison, evaluation, decision-making, refinement, and justification. Four of these stages, translation, AI consultation, refinement, and written justification, were already built into the task. The portfolio analysis makes comparison, evaluation, and decision-making more explicit as intervening processes. The sequence is therefore a pedagogical formulation grounded in this bounded dataset, not a validated model of instructional effectiveness.

Acknowledgements:	The authors thank Universiti Sains Islam Malaysia for the resources and support provided for this research and the students whose anonymised portfolio materials formed the evidence base.
Funding Statement:	This research received financial support from Universiti Sains Islam Malaysia under Grant Number PPPI/PTJ/FPBU/USIM/186626. The funding body had no role in the design of the study, data collection, analysis, interpretation of results, or the decision to publish this manuscript.
Conflict of Interest Statement:	The authors declare no conflict of interest regarding the publication of this paper. All authors contributed to the work and approved the final manuscript.
Ethics Statement:	Consent was obtained from all six participants for the research use and publication of their anonymised portfolio materials. All cases were anonymised as P1-P6, and no personally identifying student information is reported.
Author Contribution Statement:	All authors contributed significantly to the development of this manuscript. Sulhah Ramli was responsible for conceptualisation,

methodology, and overall supervision. Hishomudin Ahmad contributed to data handling, analysis, and interpretation. Muhammad Izuan Abd Gani and Aishah Isahak contributed to the literature review, drafting, and critical revision. All authors read and approved the final version.

References

Angelone, E. (2010). Uncertainty, uncertainty management and metacognitive problem solving in the translation task. In G. M. Shreve & E. Angelone (Eds.), Translation and cognition (pp. 17–40). John Benjamins. https://doi.org/10.1075/ata.xv.03ang

Boase-Beier, J. (2020). Translation and style (2nd ed.). Routledge. https://doi.org/10.4324/9780429327322

Eskelinen, J., & Pakkala-Weckström, M. (2016). Assessing translation students' acquisition of professional competences. Translation Spaces, 5(2), 314–331. https://doi.org/10.1075/ts.5.2.08esk

González Davies, M. (2004). Multiple voices in the translation classroom: Activities, tasks and projects. John Benjamins. https://doi.org/10.1075/btl.54

Hurtado Albir, A. (2015). The acquisition of translation competence: Competences, tasks, and assessment in translator training. Meta, 60(2), 256–280. https://doi.org/10.7202/1032857ar

Jiménez-Crespo, M. A. (2025). 'If students translate like a robot …' or how research on human-centered AI and intelligence augmentation can help realign translation education. The Interpreter and Translator Trainer, 19(3–4), 277–295. https://doi.org/10.1080/1750399X.2025.2542022

Kenny, D., & Winters, M. (2020). Machine translation, ethics and the literary translator's voice. Translation Spaces, 9(1), 123–149. https://doi.org/10.1075/ts.00024.ken

Kiraly, D. C. (2000). A social constructivist approach to translator education: Empowerment from theory to practice. St. Jerome Publishing.

Lahiani, R. (2025). Beyond words: Translating nonverbal communication in Arabic poetry: A cultural turn approach. Cogent Arts & Humanities, 12(1), Article 2526941. https://doi.org/10.1080/23311983.2025.2526941

Massey, G., & Ehrensberger-Dow, M. (2011). Commenting on translation: Implications for translator training. The Journal of Specialised Translation, 16, 26–41. https://doi.org/10.26034/cm.jostrans.2011.485

Moorkens, J. (2020). 'A tiny cog in a large machine': Digital Taylorism in the translation industry. Translation Spaces, 9(1), 12–34. https://doi.org/10.1075/ts.00019.moo

O'Brien, S. (2012). Translation as human-computer interaction. Translation Spaces, 1(1), 101–122. https://doi.org/10.1075/ts.1.05obr

PACTE Group, Hurtado Albir, A., Galán-Mañas, A., Kuznik, A., Olalla-Soler, C., Rodríguez-Inés, P., & Romero, L. (2018). Competence levels in translation: Working towards a European framework. The Interpreter and Translator Trainer, 12(2), 111–131. https://doi.org/10.1080/1750399X.2018.1466093

Pietrzak, P. (2019). Scaffolding student self-reflection in translator training. Translation and Interpreting Studies, 14(3), 416–436. https://doi.org/10.1075/tis.18029.pie

Risku, H. (2002). Situatedness in translation studies. Cognitive Systems Research, 3(3), 523–533. https://doi.org/10.1016/S1389-0417(02)00055-4

Romaniuk-Cholewska, D. (2025). Analysis of critical thinking skills: How do students perceive and evaluate AI translation tools? Linguodidactica, 29, 173–187. https://doi.org/10.15290/lingdid.2025.29.12

Schön, D. A. (1983). The reflective practitioner: How professionals think in action. Basic Books.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. Mind & Language, 25(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Xu, S., Su, Y., & Liu, K. (2025). Investigating student engagement with AI-driven feedback in translation revision: A mixed-methods study. Education and Information Technologies, 30(12), 16969–16995. https://doi.org/10.1007/s10639-025-13457-0

Zhang, W., Li, A. W., & Wu, C. (2025). University students' perceptions of using generative AI in translation practices. Instructional Science, 53(4), 633–655. https://doi.org/10.1007/s11251-025-09705-y