Assessing English-Arabic Medical Translation Performance among Undergraduate Students at the Faculty of Languages-Surman, Sabratha University

أداء طلاب قسم اللغة الإنجليزية في الترجمة الطبية من الإنجليزية إلى العربية: تقييم قائم على الأداء في كلية اللغات صرمان، جامعة صبراتة، ليبيا

مها عبد الله سعيد الشوشان

جامعة الزنتان /كلية التربية يفرن / قسم اللغة الإنجليزية

الدرجة العلمية: ماجستير ترجمة فورية وتحريرية

maha .alshoshan 1982 @gmail .com

maha. alshoshan @uoz,edu.ly

0926776809

ARTICLE INFO

Received: 26-04-2026

Accepted: 15-05-2026

Published: 02-06-2026

Keywords: medical translation; translation competence; performance assessment; English-Arabic translation; translator education; Libya

ABSTRACT

Medical translation requires strong language skills, accurate knowledge of medical terminology, and the ability to convey specialized information clearly and appropriately to the target audience. However, performance-based evidence on English-Arabic medical translation among Libyan undergraduate students remains limited.

This study assessed the performance of English Department students at the Faculty of Languages-Surman, Sabratha University, in translating a medical text from English into Arabic. Students from different semester levels completed a controlled patient-oriented medical translation task and a short questionnaire concerning their previous translation experience, use of translation resources, and perceived difficulties. The translations were independently assessed by two trained evaluators using an analytic scoring rubric covering meaning accuracy, medical terminology, Arabic-language quality, audience appropriateness, and completeness and consistency.

The findings showed an overall moderate level of translation performance. Meaning accuracy was the strongest area, while medical terminology represented

the main area of weakness. Performance also differed across semester groups, suggesting that current course direction and translation experience may influence students' ability to complete English-Arabic medical translation tasks. Terminology errors were the most common problems identified in the translated texts, followed by Arabic-language errors and unjustified omissions or additions. Students also frequently reported difficulty in choosing among alternative Arabic equivalents for medical terms.

Previous study of specialized translation was associated with better translation performance, whereas frequent use of translation resources alone did not independently indicate stronger performance. Overall, the findings highlight the need for greater attention to medical terminology, clear patient-oriented Arabic, critical evaluation of translation resources, and systematic revision in undergraduate translation training.

الملخص:

هدفت هذه الدراسة إلى تقييم أداء طلاب قسم اللغة الإنجليزية بكلية اللغات صرمان، جامعة صبراتة، في ترجمة نص طبي من الإنجليزية إلى العربية، حيث شارك في الدراسة طلاب من فصول دراسية مختلفة، فقد ترجم الطلاب نصًا طبيًا قصيرًا موجهًا للمرضى، ثم أجابوا عن استبيان قصير حول خبرتهم السابقة في الترجمة، والمصادر التي يستخدمونها، والصعوبات التي يواجهونها. بعد ذلك، قيّم مقيمان مدربان ترجمات الطلاب بصورة مستقلة، وفق معايير شملت دقة نقل المعنى، وصحة المصطلحات الطبية، وسلامة اللغة العربية، ووضوح الأسلوب، واكتمال الترجمة واتساقها.

أظهرت النتائج أن مستوى الطلاب في الترجمة كان متوسطًا بشكل عام، وكان أداؤهم أفضل في نقل المعنى، بينما واجهوا صعوبة أكبر في استخدام المصطلحات الطبية، كما اختلف مستوى الأداء بين الفصول الدراسية، وقد يكون ذلك مرتبطًا بنوع مقرر الترجمة والخبرة السابقة لدى الطلاب، وكانت أخطاء المصطلحات الطبية الأكثر شيوعًا، تلتها أخطاء اللغة العربية، ثم الحذف أو الإضافة، كذلك وجد كثير من الطلاب صعوبة في اختيار المقابل العربي المناسب للمصطلحات الطبية عند وجود أكثر من ترجمة ممكنة.

كما حقق الطلاب الذين سبق لهم دراسة الترجمة المتخصصة أداءً أفضل، بينما لم يكن كثرة استخدام مصادر الترجمة وحدها مرتبطة بتحسن الأداء، وتشير هذه النتائج إلى أهمية تدريب الطلاب على اختيار المصطلحات الطبية الصحيحة، وكتابة نص عربي واضح ومناسب للمرضى، واستخدام مصادر الترجمة بعناية، ومراجعة النصوص المترجمة قبل اعتمادها.

الكلمات المفتاحية: الترجمة الطبية؛ كفاءة الترجمة؛ التقييم القائم على الأداء؛ الترجمة من الإنجليزية إلى العربية؛ إعداد المترجمين؛ ليبيا

Introduction

Medical translation mediates information that may influence clinical decisions, medication use, informed consent, and public understanding of health. Its quality therefore depends not only on linguistic fluency but also on terminological precision, accurate representation of risk, control of negation and quantity, audience-sensitive register, and rigorous revision. A translation can be grammatically acceptable while remaining clinically misleading if it suppresses a qualifier, selects an inappropriate equivalent, or obscures the force of an instruction. Recent evaluations of multilingual health information have likewise shown that apparently fluent output may contain terminological and syntactic problems that require expert human review (Delfani *et al.*, 2024; Rao *et al.*, 2024).

English-Arabic medical translation presents additional challenges arising from structural differences between the two languages and from variation in Arabic medical terminology. English medical discourse frequently uses dense noun phrases, passive structures, abbreviations, Greek- and Latin-derived formations, and compact relations between modifiers and heads. Arabic may offer a recognized equivalent, an Arabized form, a descriptive rendering, or several competing terms for the same concept. Consequently, successful translation requires concept identification and contextual verification rather than word substitution (Al-Jarf, 2018; Kovács, 2023; Mahdi & Mohammed, 2025).

This issue is particularly relevant in Libya, where English is widely used in medical education and professional documentation, while Arabic is essential for communication with patients and the wider public. Aleskandarani (2023), working at the University of Benghazi, found substantial inaccuracies when medical terms and abbreviations were translated from English into Arabic by widely used machine-translation systems. The study attributed recurring problems to literal rendering, transliteration, missing equivalents, and inadequate contextual

interpretation. These findings underline the importance of developing human translators who can evaluate terminology and revise technology-assisted output rather than accept it uncritically.

Research on Libyan translation students also identifies persistent performance gaps. Alsunousi (2025) found recurrent grammatical, lexical, omission, coherence, and register problems in English-Arabic translations produced by undergraduates at Misurata University. Rafah and Elraggas (2024) similarly reported Arabic grammatical errors in translations produced by Libyan graduate translation students, despite their ability to infer some meanings from context. More recently, Saleh (2026) found that Libyan undergraduates translating scientific texts frequently relied on transference, naturalization, and literal translation and used recognized or functional equivalents less consistently. Together, these studies suggest that general language study and exposure to translation courses do not automatically produce competence in specialized English-Arabic translation.

Translation competence is now widely conceptualized as a coordinated set of abilities. The European Master's in Translation framework includes language and culture, translation, technology, personal and interpersonal, and service-provision competence (European Commission, 2022). The PACTE (Process in the Acquisition of Translation Competence and Evaluation) framework similarly emphasizes bilingual, extralinguistic, instrumental, strategic, and knowledge-about-translation components (Hurtado Albir, 2017). In specialized medical translation, thematic knowledge, information retrieval, target-language production, and strategic monitoring become particularly salient. A performance-based task is therefore better suited than self-report alone to determine whether students can mobilize these abilities in a complete text.

At Sabratha University, translation is taught across different semesters and language directions. Semester Five students study translation from English into Arabic, Semester Six students study translation from Arabic into English, and Semester Eight students study Selective Translation. These courses expose students to different translation directions, text types, linguistic structures, and problem-solving demands. Medical translation represents one context in which students are required to coordinate source-text comprehension, terminological selection, target-language production, and attention to the intended audience when translating from English into Arabic.

Problem Statement

Undergraduate English majors may encounter difficulty when translating medical texts from English into Arabic because this type of translation requires the simultaneous application of linguistic, terminological, thematic, and strategic abilities. Students may understand the general meaning of a source text but still experience problems in selecting accurate medical equivalents, interpreting abbreviations, restructuring complex English expressions, producing natural Modern Standard Arabic, and adapting the target text to the needs of patients and other non-specialist readers.

These difficulties may arise from limited medical background knowledge, variation in Arabic medical terminology, dependence on literal translation, inadequate evaluation of dictionaries and digital resources, and insufficient practice in translating complete medical texts. Reliance on unverified machine-translation output may further increase terminological inconsistency and reduce the accuracy of the translated text.

Such problems may result in distortion of meaning, inappropriate terminology, omission or addition of clinically relevant information, unclear Arabic constructions, inconsistent rendering of repeated concepts, and unsuitable register. In patient-oriented texts, these weaknesses may affect the clarity, completeness, and functional adequacy of health information. The problem addressed in this study is therefore the ability of English Department students to integrate meaning accuracy, medical terminology, Arabic-language quality, audience-appropriate style, and textual completeness when translating a medical patient-information text from English into Arabic.

Significance of the Study

This study may contribute locally grounded evidence concerning English-Arabic medical-translation performance among undergraduate English majors. By assessing complete student translations rather than relying solely on self-reported confidence, it may provide a more direct indication of students' ability to translate a medical patient-information text. The analytic rubric may also help distinguish performance across meaning accuracy, medical terminology, Arabic-language quality, audience-appropriate style, and textual completeness and consistency.

The findings may assist the English Department in identifying areas that could require greater attention within translation courses. They may inform decisions concerning course sequencing, terminology training, Arabic writing practice, resource evaluation, revision activities, and performance-based assessment. The study may also provide a basis for subsequent research involving other medical genres, larger student populations, or additional

Libyan institutions. More broadly, it could contribute to developing assessment practices that reflect the linguistic, terminological, and functional demands of specialized English-Arabic translation.

Objectives of the Study

The study aims to assess students' performance in translating a medical text from English into Arabic. Specifically, it seeks to:

1. Assess students' overall medical translation performance across the main assessment components and determine whether performance differs across semester groups.
2. Identify the most common translation errors and difficulties reported by students.
3. Examine whether previous specialized-translation study and resource use are associated with translation performance.

Research Questions

1. What is the overall level of students' English-Arabic medical translation performance across the assessed components, and does performance differ across semester groups?
2. What are the most common translation errors and difficulties reported by students?
3. Is previous specialized-translation study and resource use associated with students' translation performance?

Hypotheses

The study assumes that students who have previously studied specialized translation are likely to achieve better medical translation performance. It also assumes that the use of translation resources may be associated with students' translation performance.

Literature Review

1. Medical Translation, Terminology, and Audience

Research on medical translation converges on three interconnected requirements: conceptual accuracy, terminological control, and adaptation to the intended reader. Medical genres range from research and clinical documentation to consent materials and patient information; consequently, an equivalent that is acceptable in specialist communication may be unsuitable or incomprehensible in patient-facing Arabic. The risk is especially pronounced in instructions involving dosage, negation, temporal sequence, allergy, severity, and urgent action. Studies of medical and health translation therefore treat terminology and audience design as components of

meaning rather than as optional stylistic refinements (Al-Jarf, 2018; Kovács, 2023; Mahdi & Mohammed, 2025; Rao *et al.*, 2024; Strasly, 2026).

Within English-Arabic translation, the evidence consistently identifies competing Arabic equivalents, insufficient standardization, abbreviation ambiguity, transliteration, and literal rendering as recurrent sources of error. These problems are related: the absence of a single stable equivalent encourages translators to depend on the first dictionary or machine-generated option, while inadequate contextual verification increases inconsistency within the same target text. Accordingly, competent performance requires concept identification, comparison of authoritative sources, and consistent use of an equivalent suited to both the medical concept and the target audience (Al-Jarf, 2018; Aleskandarani, 2023; Mahdi & Mohammed, 2025).

2. Performance Assessment and Recurrent Student Errors

Translation-competence frameworks conceptualize performance as the coordinated use of bilingual, extralinguistic, instrumental, strategic, and translation-specific knowledge (European Commission, 2022; Hurtado Albir, 2017). Research on specialized translation and metacognition likewise indicates that subject knowledge, strategic monitoring, and students' perceptions do not represent interchangeable dimensions of competence (Huang, 2023; Xie & Wang, 2025). This multidimensional view has direct assessment implications. Confidence questionnaires can identify perceived difficulty, but they cannot establish whether a student preserved a negative instruction, selected an accurate term, or produced readable Arabic. Whole-text tasks scored with explicit analytic criteria provide more direct evidence and allow meaning, terminology, target-language quality, audience, and completeness to be interpreted separately (Man *et al.*, 2022; Salamah, 2021).

Empirical error research reinforces this component-based approach. Across Arabic translation settings, recurring weaknesses cluster around inaccurate meaning transfer, lexical selection, grammar, omission, cohesion, and register rather than a single isolated skill, and recent regional evidence points to continuing linguistic and intercultural gaps in translator preparation (Mekheimer, 2026). Verified Libyan evidence shows the same pattern. In translations produced by ten Misurata University undergraduates, Alsunousi (2025) recorded 72 language errors, 68 rendition errors, and 13 omissions; grammatical errors were the largest individual category with 38 occurrences. Rafah and Elraggas (2024) likewise reported contextual and target-language difficulties among Libyan graduate translation students, while Saleh (2026) identified inaccurate transference and naturalization in scientific

translation. Taken together, these findings justify assessment that combines component scores with coded error patterns.

3. Technology, Resources, and Instrumental Competence

Technology research distinguishes access or frequency of use from instrumental competence. Effective use of dictionaries, terminology databases, corpora, machine translation, and generative systems requires source selection, comparison, confidentiality awareness, consistency management, and critical post-editing. Translation-technology literacy therefore concerns the quality of decisions made with tools, not simply how often a tool is used, and its development involves both student and educator preparedness (Bowker, 2020; Bowker & Ciro, 2019; Cai *et al.*, 2025; Hackenbuchner & Krüger, 2023; Hazaea & Qassem, 2024; Krüger, 2022, 2023, 2024; Zhang & Doherty, 2025).

Medical translation provides a particularly clear test of that distinction. Aleskandarani (2023) compared Google, Bing, and Yandex across 45 items divided equally into 15 COVID-19 terms, 15 other medical terms, and 15 medical abbreviations. The outputs documented recurrent problems involving literal rendering, transliteration, missing equivalents, and inadequate contextual interpretation across the three systems. Broader reviews similarly emphasize the continuing difficulty of producing contextually appropriate automated translations (Naveen & Trojovský, 2024). This evidence confirms that apparently fluent automated wording cannot be accepted without terminological verification and supports evaluating students' ability to justify and revise technology-assisted choices.

4. Libyan Translator Education and the Present Research Gap

Libyan studies connect individual translation problems to wider curricular and institutional conditions. In a mixed-method investigation involving 150 students and four translation teachers, AL-Darraji (2023) found unanimous agreement in the relevant survey theme that restricted access to materials, technology, and funding affected translation education; 93% also agreed that inadequate infrastructure and support systems affected delivery. Performance studies from Misurata and other Libyan settings simultaneously document persistent grammatical, lexical, coherence, register, and specialized-terminology problems (Alsunousi, 2025; Rafah & Elraggas, 2024; Saleh, 2026). The combined evidence suggests that specialized performance reflects both learner competence and the opportunities, resources, and feedback structures available in translation programs.

Despite this growing evidence, the Libyan literature remains fragmented across perceptions of program constraints, analyses of general or scientific translation errors, and evaluations of automated terminology. A common performance task assessing a complete English-Arabic medical text across semester groups has not been reported for the Faculty of Languages-Surman. The present study addresses that gap by combining an analytic performance rubric, independent scoring, categorical error analysis, and a supporting questionnaire. It therefore examines not only whether students experience difficulty, but where that difficulty appears in an observable target text and whether previous specialized-translation study retains an association with performance after semester is considered.

Methodology

Research Design

The study used a quantitative cross-sectional performance-assessment design. A controlled translation task provided the primary outcome, while a short questionnaire supplied contextual information on previous study, resource use, and perceived difficulty. The design permitted descriptive profiling and group comparisons but did not establish causal development across semesters.

Setting and Participants

The study was conducted in the Department of English, Faculty of Languages-Surman, Sabratha University, Libya, during Spring 2025-2026. The accessible population comprised 51 students: 10 in Semester Five, 28 in Semester Six, and 13 in Semester Eight. At the time of data collection, Semester Five students were taking Translation from English into Arabic, Semester Six students were taking Translation from Arabic into English, and Semester Eight students were taking Selective Translation as presented in Table 1. A census approach was used, and all eligible students were invited. All 51 students completed both instruments, yielding a response rate of 100.0%.

Table 1. Participant distribution by semester and current translation course

Semester	Current course	Eligible n	Participating n
Five	Translation from English into Arabic	10	10
Six	Translation from Arabic into English	28	28
Eight	Selective Translation	13	13
Total	-	51	51

Translation Task

Participants translated a 170-word patient-information text on safe antibiotic use from English into Arabic. The passage was selected because it combines technical and semi-technical terminology with patient-directed instructions, negation, conditional meaning, risk statements, and repeated concepts requiring consistency. It includes ten pre-identified terminology units: bacterial infection, viral infection, prescribed course, antimicrobial resistance, allergic reaction, urgent medical help, severe skin rash, persistent diarrhea, drug interaction, and missed dose. The target audience is defined as adult Arabic-speaking patients and their families.

All participants received the same text, instructions, and 45-minute time limit. Under the controlled-resource condition, printed and electronic dictionaries and terminology databases were permitted, whereas machine-translation and generative-AI systems were not permitted. Student names were replaced with anonymous codes before scoring. The test appears in Appendix A.

Analytic Scoring Rubric

The rubric allocated 100 points across meaning accuracy (30), medical terminology (25), Arabic language quality (20), style and audience appropriateness (15), and completeness and consistency (10). Criterion descriptors distinguished minor problems from errors that altered clinical meaning. Two raters with English-Arabic translation expertise scored each script independently. The final score was the mean of the two ratings unless the ratings differed by more than 10 points, in which case the script was reviewed through adjudication. Appendix B presents the full rubric.

Questionnaire

The questionnaire did not measure competence. It recorded semester, previous study of specialized or medical translation, frequency and type of resource use, and difficulties experienced during medical translation. Difficulty variables were coded independently because participants could select more than one option. Anonymous codes linked questionnaire responses to test scores and permitted analysis of whether previous study and reported practices corresponded to observed performance. Appendix C contains the questionnaire.

Validity and Reliability

Five specialists in English-Arabic translation, medical language, language assessment, and research methods evaluated the task, instructions, terminology list, rubric, and questionnaire for relevance, clarity, representativeness, and difficulty, following established principles of instrument development and content validation (DeVellis &

Thorpe, 2021). Item-level content-validity indices ranged from 0.80 to 1.00, and the average scale-level content-validity index was 0.94. The instruments were piloted with eight English Department students who were not included in the analytic sample. The pilot confirmed the 45-minute time limit and led to minor clarification of two questionnaire response options and one rubric descriptor; the medical text and scoring weights were unchanged.

The raters completed a calibration session using practice scripts. Inter-rater reliability for continuous total and component scores was estimated using a two-way random-effects, absolute-agreement, single-measure intraclass correlation coefficient. The ICC model, unit, and confidence interval were reported because a coefficient without these specifications is ambiguous (Koo & Li, 2016). Agreement was also examined through the mean rater difference and the number of scripts requiring adjudication.

Data Collection Procedure

After institutional approval, eligible students received written information describing the study, voluntary participation, confidentiality, and the separation of research scores from course grades. Consenting participants completed the coded questionnaire and translation task under standardized conditions. Scripts were checked for completeness, anonymized, and distributed to both raters in different random orders. Questionnaire data, rater scores, and coded error frequencies were entered into a verification file and independently checked against the original forms.

Ethical Considerations

Ethical approval was obtained from the Faculty of Languages-Surman, Sabratha University, prior to data collection (Approval No. 164). All eligible students received written information explaining the purpose of the study, the voluntary nature of participation, confidentiality, and the separation of research scores from course grades. Informed consent was obtained from all participants before completion of the questionnaire and translation task. Anonymous codes were used in place of student names to protect participants' identities, and all collected data were handled confidentially.

Error Analysis

In addition to the rubric scores, translation errors were coded into five categories: meaning distortion, terminology error, unjustified omission or addition, Arabic-language error, and register or cohesion problem. A single segment could receive more than one code only when the identified problems were analytically distinct. A

coding manual defined each category and provided illustrative examples. Fifteen randomly selected scripts, representing 29.4% of the sample, were independently coded by both raters. Inter-coder agreement was evaluated using Cohen's kappa, and disagreements were resolved through discussion before the full frequency analysis was completed.

Data Analysis

Data were inserted and analyzed by using SPSS version 26.0. Categorical variables were summarized using frequencies and percentages. Continuous test and component scores were reported as means, standard deviations, medians, interquartile ranges, ranges, and 95% confidence intervals. Distributional shape, influential observations, and variance equality were inspected before group comparison. Because the semester groups were unequal and homogeneity could not be assumed, total scores were compared using Welch's one-way ANOVA, followed by Games-Howell pairwise comparisons. Effect magnitude was summarized using omega squared with a 95% confidence interval.

Previous specialized-translation study was first examined using Welch's independent-samples t test, with Hedges' g and its 95% confidence interval. Tool-use frequency was coded as Rarely = 1, Sometimes = 2, and Often = 3. The bivariate association between ordinal tool-use frequency and total scores was assessed using Spearman's rho. A multiple linear regression model was then fitted to estimate the independent contributions of previous specialized-translation study and tool-use frequency while adjusting for semester. Semester Six, the largest group, was used as the reference category. Linearity was inspected using partial-residual plots; homoscedasticity was assessed using the residual-versus-fitted plot and the Breusch–Pagan test; approximate residual normality was assessed using the Q–Q plot and Shapiro–Wilk test; multicollinearity was evaluated using variance-inflation factors; and influential observations were examined using standardized residuals and Cook's distance, in accordance with established multivariate-analysis guidance. Unstandardized coefficients, standard errors, standardized coefficients, exact p values, model R^2, and adjusted R^2 were reported. Exact p values were reported to three decimal places, except $p < 0.001$. Statistical conclusions were based on p values, effect sizes, confidence intervals, and the precision permitted by the small population rather than on statistical significance alone.

Results

Participant Characteristics

Of the 51 eligible students, 51 completed both instruments, giving a response rate of 100.0%. The final sample included 10 students from Semester Five, 28 from Semester Six, and 13 from Semester Eight. Overall, 21 students (41.2%) reported previous study of specialized or medical translation. Participant characteristics and resource-use frequency are summarized in Table 2. The most frequently used resources were general web search (78.4%), followed by machine translation (64.7%).

Table 2. Participant characteristics

Variable	Category	n	%
Semester	Five	10	19.6
	Six	28	54.9
	Eight	13	25.5
Previous specialized/medical translation	Yes	21	41.2
	No	30	58.8
Resource use	Rarely	12	23.5
	Sometimes	25	49.0
	Often	14	27.5

Inter-Rater Reliability

Table 3 presents the inter-rater reliability results. Agreement between the two raters on the total translation-test scores was excellent, ICC(2,1; absolute agreement) = 0.94, 95% CI [0.90, 0.97]. The mean difference between the raters was 0.6 points, and three scripts required adjudication. Component-level ICC values ranged from 0.84 to 0.94. Inter-coder agreement for the translation-error categories was also strong. Based on the fifteen scripts that were independently coded by both raters, Cohen's κ was 0.86, 95% CI [0.80, 0.92].

Table 3. Inter-rater reliability for translation scores

Outcome	ICC	95% CI	Mean rater difference
Total score	0.94	[0.90, 0.97]	0.6
Meaning accuracy	0.91	[0.85, 0.95]	0.2
Terminology	0.89	[0.81, 0.94]	0.1
Arabic language	0.87	[0.78, 0.93]	0.1
Style/register	0.84	[0.72, 0.91]	0.1
Completeness/consistency	0.86	[0.76, 0.92]	0.1

Note. ICC estimates were calculated using a two-way random-effects, absolute-agreement, single-measure model.

Overall and Component Performance

The mean total score was 67.8 (SD = 10.9; median = 68.0, IQR = 15.0; range = 44-89). Among rubric components, meaning accuracy had the highest percentage of its maximum score (73.3%), whereas medical terminology had the lowest (61.6%) as illustrated in Table 4.

Table 4. Overall and component-level translation performance

Outcome	Maximum	Mean	SD	Median (IQR)	Mean % of maximum
Meaning accuracy	30	22.0	3.8	22.0 (5.0)	73.3
Medical terminology	25	15.4	3.7	15.0 (5.0)	61.6
Arabic language quality	20	13.9	2.8	14.0 (4.0)	69.5
Style/audience appropriateness	15	10.1	2.2	10.0 (3.0)	67.3
Completeness/consistency	10	6.4	1.7	6.0 (2.0)	64.0
Total	100	67.8	10.9	68.0 (15.0)	67.8

Differences across Semesters

Table 5 reveals differences in translation performance across semesters. The mean total scores were 74.6 (SD = 8.1) in Semester Five, 63.9 (SD = 10.4) in Semester Six, and 70.9 (SD = 10.2) in Semester Eight. The omnibus comparison was statistically significant, Welch's $F = 5.36$, $df = 2, 23.7$, $p = 0.012$, effect size = omega squared = 0.15, 95% CI [0.02, 0.29]. Post-hoc analysis showed that Semester Five outperformed Semester Six (mean difference = 10.7, $p = 0.009$), whereas the other contrasts were not statistically significant ($ps > 0.100$).

Table 5. Translation performance by semester

Semester	n	Mean	SD	95% CI	Median (IQR)
Five	10	74.6	8.1	[68.8, 80.4]	75.0 (11.0)
Six	28	63.9	10.4	[59.9, 67.9]	64.0 (14.0)
Eight	13	70.9	10.2	[64.7, 77.1]	71.0 (15.0)

Error Patterns and Reported Difficulties

The most frequent error category was terminology errors (112 occurrences; 33.8%), followed by Arabic language errors (73; 22.1%) and unjustified omissions or additions (55; 16.6%). At segment level, recurrent problems involved literal or inconsistent equivalents for antimicrobial resistance, failure to preserve the urgency encoded in seek urgent medical help, and awkward Arabic clause linkage that obscured conditional instructions. In the questionnaire, the most commonly selected difficulty was multiple Arabic equivalents (37, 72.5%), followed by

medical abbreviations (35, 68.6%). Agreement for the double-coded subset was strong (Cohen's kappa = 0.86, 95% CI [0.80, 0.92]). The distributions of error categories and reported difficulties are presented in Table 6.

Table 6. Translation errors and student-reported difficulties

Indicator	n/occurrences	% of students or errors
Meaning distortion	48	14.5
Terminology error	112	33.8
Omission/addition	55	16.6
Arabic language error	73	22.1
Register/cohesion problem	43	13.0
Multiple Arabic equivalents reported as difficult	37	72.5
Medical abbreviations reported as difficult	35	68.6
Insufficient medical background reported	32	62.7

Note. Percentages for the five error categories use the 331 coded errors as the denominator; percentages for reported difficulties use the 51 participants as the denominator. Multiple difficulty responses were permitted.

Factors Associated with Performance

Table 7 shows the factors associated with performance. Students with previous specialized-translation study obtained a mean score of 72.5 (SD = 9.1) compared with 64.5 (SD = 10.8) among those without such study. The unadjusted difference was statistically significant, Welch's $t(47.2) = 2.86$, $p = 0.006$, Hedges' $g = 0.78$, 95% CI [0.20, 1.35]. Tool-use frequency was positively associated with total score at the bivariate level, Spearman's rho = 0.31, $p = 0.028$. In the adjusted regression model, the four predictors explained 29.0% of the variance in total scores, $R^2 = 0.29$, adjusted $R^2 = 0.23$, $F(4, 46) = 4.70$, $p = 0.003$. Previous specialized-translation study remained independently associated with higher performance ($B = 5.6$, $SE = 2.6$, $\beta = 0.25$, $p = 0.037$), as did membership in Semester Five relative to Semester Six ($B = 8.2$, $SE = 3.5$, $\beta = .30$, $p = 0.023$). The Semester Eight contrast ($B = 4.5$, $SE = 3.0$, $\beta = .18$, $p = 0.141$) and tool-use frequency ($B = 2.1$, $SE = 1.5$, $\beta = .16$, $p = 0.168$) were not independently significant.

Regression diagnostics supported the model assumptions. Partial-residual plots were approximately linear, and the residual-versus-fitted plot showed no systematic pattern. The Breusch-Pagan test did not indicate heteroscedasticity, $\chi^2(4) = 4.11$, $p = 0.391$. Residuals were approximately normally distributed, Shapiro-Wilk $W = 0.98$, $p = 0.573$. Standardized residuals ranged from -2.41 to 2.52, and the maximum Cook's distance was 0.18; thus,

no observation exceeded the conventional Cook's distance threshold of 1.00. Variance-inflation factors ranged from 1.08 to 1.34, indicating no problematic multicollinearity.

Table 7. Multiple linear regression predicting total translation-test score

Predictor	B	SE	β	p	95% CI for B
Semester Five vs. Six	8.2	3.5	0.30	0.023	[1.15, 15.25]
Semester Eight vs. Six	4.5	3.0	0.18	0.141	[-1.54, 10.54]
Previous specialized translation	5.6	2.6	0.25	0.037	[0.37, 10.83]
Tool-use frequency	2.1	1.5	0.16	0.168	[-0.92, 5.12]

Note. N = 51. Semester Six was the reference category. Model $R^2 = 0.29$; adjusted $R^2 = 0.23$; $F(4, 46) = 4.70$, $p = 0.003$. B = unstandardized coefficient; β = standardized coefficient; CI = confidence interval.

Discussion

This study provides a performance-based account of English-Arabic medical translation among English majors at Sabratha University. The findings indicate a moderate overall level of performance accompanied by clear variation among students. This variation suggests that completion of general language and translation courses does not necessarily result in comparable control of all the abilities required for medical translation. Students may possess sufficient general comprehension to identify the principal message of a medical text while differing considerably in their ability to manage specialized terminology, produce natural Arabic, preserve clinically relevant details, and adapt the translation to a non-specialist audience. This interpretation is consistent with the multidimensional view of translation competence proposed by the Hurtado Albir (2017) and the European Commission (2022), both of which treat translation as the coordinated use of linguistic, thematic, instrumental, and strategic abilities rather than as an extension of bilingual proficiency alone.

Meaning accuracy emerged as the comparatively strongest component, whereas medical terminology represented the principal area of weakness. One explanation is that the patient-information text conveyed a recognizable communicative purpose and consisted largely of instructions and warnings whose general meanings could be inferred from context. Students may therefore have understood the main propositions even when they lacked precise Arabic equivalents for specialized concepts. Terminological selection, by contrast, requires concept-specific knowledge and discrimination among alternatives that may appear linguistically acceptable but differ in medical precision or communicative suitability. The difference between meaning preservation and terminological

accuracy therefore suggests that students' general bilingual resources were more developed than their specialized terminological competence.

This finding agrees with Al-Jarf (2018), who identified variation among Arabic equivalents as a major source of difficulty in English-Arabic medical translation. It also corresponds with Mahdi and Mohammed (2025), whose findings demonstrated continuing problems in the accurate rendering of medical terminology. The present study extends these findings by showing that terminological difficulty is visible not only in isolated term translation but also within a complete patient-information text, where the selected equivalent must remain accurate, readable, and consistent across the translation. Unlike studies based exclusively on individual terms, the current performance task demonstrates how terminological decisions interact with syntax, audience, cohesion, and preservation of meaning.

The error analysis and questionnaire responses converged in identifying terminology as the most prominent difficulty. This convergence is important because the two instruments capture different forms of evidence. Error coding reflects difficulties that appeared in students' actual translations, whereas questionnaire responses reflect the difficulties students recognized and reported themselves. Their agreement strengthens the interpretation that terminological uncertainty was a genuine feature of performance rather than an artefact of the scoring procedure. It also suggests that students were at least partly aware of their difficulty in choosing among alternative Arabic equivalents. However, awareness of the problem did not necessarily provide the strategies required to resolve it.

The result is also compatible with Aleskandarani's (2023) analysis of English-Arabic medical terminology produced by general-purpose machine-translation systems. That study documented recurring problems involving literal rendering, transliteration, missing equivalents, and insufficient contextual interpretation. Similar tendencies in students' translations may reflect reliance on readily available equivalents without adequate verification. The present study does not demonstrate that students copied automated output, particularly because such systems were not permitted during the task. Rather, the similarity indicates that both novice translators and automated systems may produce superficially plausible solutions when contextual and conceptual evaluation is insufficient. The finding therefore reinforces the distinction between obtaining a proposed equivalent and determining whether that equivalent is medically accurate and communicatively appropriate.

Arabic-language errors constituted another important constraint on performance. Students sometimes appeared to understand the English source but encountered difficulty reconstructing its meaning in clear and

natural Modern Standard Arabic. Medical patient-information texts require a balance between accuracy and accessibility. Excessively literal restructuring may preserve the visible form of the English sentence while producing awkward Arabic syntax, unclear reference, weak cohesion, or ambiguity in conditional and urgent instructions. Conversely, excessive reformulation may improve fluency but omit qualifications or alter the strength of medical advice. The observed Arabic-language problems can therefore be interpreted as evidence of difficulty coordinating source-text accuracy with target-language naturalness.

This pattern closely corresponds with Alsunousi's (2025) findings among Libyan undergraduates, which identified grammatical inaccuracies, lexical omission, weak coherence, and register problems in English-Arabic translation. It also agrees with Rafah and Elraggas (2024), who reported Arabic grammatical difficulties among Libyan graduate translation students even when contextual clues supported comprehension of the source expression. Together, these studies indicate that target-language weakness may persist across educational levels and translation genres. The present study adds a medical and patient-oriented dimension to this evidence by showing that Arabic-language problems may affect the communication of instructions, warnings, and health-related qualifications. In this context, target-language errors are not merely stylistic defects; they may alter readability, emphasis, and the practical interpretation of the translated information.

The findings also correspond with Saleh's (2026) analysis of scientific translation by Libyan students. Saleh identified reliance on transference, naturalization, and literal translation when students encountered specialized terminology. The current findings suggest a related pattern in medical translation, where difficulty in locating an established equivalent may lead students to preserve the English term, reproduce its form, or construct a literal Arabic expression. The present study nevertheless extends Saleh's work by evaluating these choices within a complete text and by distinguishing terminological weakness from problems of meaning, Arabic-language quality, audience orientation, and completeness.

Differences across semester groups were statistically meaningful, but they should not be interpreted as evidence of straightforward development from one semester to another. The groups differed not only in academic position but also in the direction and scope of their current translation courses. Students studying translation from English into Arabic were working in the same direction as the research task, whereas another group was receiving current instruction in the opposite direction. Direction-matched practice may have increased familiarity with

restructuring English constructions, selecting Arabic equivalents, and revising Arabic target texts. The observed group differences may therefore reflect the relevance and recency of course experience rather than semester level alone.

This interpretation is consistent with competence-based approaches that view translation development as task-sensitive rather than purely cumulative. Hurtado Albir (2017) emphasizes that translation performance depends on the activation of several sub-competences in response to a particular task. Similarly, contextualized assessment research indicates that performance may vary according to genre, direction, audience, and available resources (Man *et al.*, 2022). The current findings consequently do not establish that students in one semester possess greater general translation ability than students in another. They show that performance on an English-Arabic medical task was associated with the curricular and directional context in which the students were studying.

Previous exposure to specialized translation was positively associated with performance, and this relationship remained evident after semester and tool-use frequency were considered. This result suggests that specialized exposure may provide benefits beyond general advancement through the program. Such experience may strengthen students' familiarity with the distinctive conceptual, textual, and communicative demands of specialized genres and may make them less likely to treat surface fluency as sufficient evidence of translation accuracy.

This finding is compatible with AL-Darraji's (2023) account of the challenges affecting translator education in Libyan universities, particularly limitations related to specialized resources, technological infrastructure, and opportunities for professional preparation. It also accords with Prieto Ramos & Guzmán's (2024) emphasis on the contribution of specialized translator training to revision performance. The present study extends these observations by showing that previous specialized-translation study remained associated with actual performance after relevant background differences were considered. Nevertheless, the cross-sectional design prevents a causal interpretation. Students who had previously studied specialized translation may also differ in motivation, academic engagement, prior language ability, or independent learning practices. Furthermore, the content and quality of their previous specialized training were not standardized.

The findings concerning resource-use frequency were more nuanced. Frequent use initially appeared to correspond with stronger performance, but it did not retain an independent association after semester and previous

specialized-translation study were considered. This suggests that frequency alone is an inadequate indicator of instrumental competence. The result may reflect differences in how students interpret and evaluate the information obtained from resources: frequent consultation does not necessarily entail stronger judgement, whereas less frequent but more deliberate use may support more accurate decisions.

This interpretation supports Bowker and Ciro's (2019) distinction between access to machine translation and informed machine-translation literacy. It is also consistent with Krüger's (2022, 2023, 2024) view that technological competence cannot be reduced to operational use and with Hazaea & Qassem's (2024) identification of limitations in instrumental competence within translator-training programs. The present study extends this literature by indicating that reported frequency of resource use, when separated from other educational factors, does not adequately explain variation in performance.

The strong agreement between the raters supports the consistency of the analytic assessment. Agreement was evident both in the assignment of continuous rubric scores and in the classification of translation errors. These two forms of reliability address different aspects of the assessment process: score agreement indicates that the raters applied the rubric similarly, while coding agreement indicates that they generally identified the same types of translation problems. This consistency reduces the likelihood that the observed performance profile resulted primarily from individual rater preferences.

The use of multiple forms of evidence also strengthens the interpretation of the findings. Rubric components identified broad areas of relative strength and weakness, error coding showed how those weaknesses appeared in the target texts, and questionnaire responses indicated whether students recognized the difficulties they encountered. This triangulated approach extends earlier Libyan studies that relied on either perceptions, selected errors, or isolated terminology. It provides a more integrated account of medical-translation performance by connecting overall achievement with specific linguistic manifestations and relevant educational experiences.

The contribution of the present study lies in connecting several forms of evidence that have often been examined separately in previous research. Earlier studies have addressed terminology problems, student translation errors, technological limitations, or institutional constraints, whereas the present analysis relates component-level performance to observed errors, perceived difficulties, curricular position, previous specialized-translation study, and resource-use practices. This integrated design provides a more differentiated account of English-Arabic medical-

translation performance and clarifies why a single total score or self-reported measure may be insufficient for identifying the sources of students' difficulties.

Conclusion

The study demonstrated that English-Arabic medical-translation performance among English majors at Sabratha University varied across the assessed components and student groups. Students were comparatively more successful in preserving the general meaning of the source text than in managing medical terminology, while target-language quality and complete transfer of information remained additional areas of difficulty.

Differences across semester groups should be interpreted in relation to course direction and curricular exposure rather than as evidence of linear academic progression. Previous specialized-translation study was positively associated with performance, whereas frequent resource use alone did not independently indicate stronger performance. Overall, the findings show that successful medical-translation performance requires the coordinated application of source-text comprehension, terminological decision-making, clear Arabic production, audience awareness, consistency, and revision.

Recommendations

1. Introduce a sequenced medical-translation unit that moves from terminology and short patient instructions to complete clinical and public-health genres.
2. Require terminology records that document the concept, contextual evidence, Arabic equivalent, source authority, and rejected alternatives.
3. Integrate advanced Modern Standard Arabic writing, cohesion, punctuation, and register into translation assessment.
4. Teach abbreviations and medical word formation through contextual tasks rather than decontextualized memorization.
5. Assess technology-assisted translation through documented source evaluation and post-editing, with explicit attention to confidentiality.
6. Use analytic rubrics and double scoring for major performance assessments, with periodic rater calibration.

7. Repeat the common task before and after targeted instruction to estimate within-student development and reduce confounding by semester.

Limitations

Several limitations define the scope of the findings. First, the study was conducted in one English Department and involved a finite population of 51 students, limiting statistical precision and external generalization. Second, the semester groups were unequal. Third, semester membership was confounded with the direction and scope of the current translation course, so group contrasts cannot be interpreted as developmental effects. Fourth, the cross-sectional design compared different students rather than tracking change within the same students over time. Fifth, one patient-information passage cannot represent the full range of medical genres or the entire construct of medical-translation competence. Sixth, the supporting questionnaire used broad self-reported categories, particularly for tool-use frequency, and did not measure the quality of resource evaluation. Error coding also involved judgment despite explicit definitions, double coding, and strong agreement. Future studies should use multiple medical genres, repeated or longitudinal measures, richer indicators of instrumental competence, and comparable course exposure across more than one institution.

References

1. AL-Darraji, O. A. O. (2023). An Exploratory Study on Improving Translation Education in Libyan Universities: A Mixed-Method Study with Student Surveys and Faculties Staff Members Interviews. *The Scientific Journal of University of Benghazi*, *36*(2). https://doi.org/10.37376/sjuob.v36i2.4291
2. Aleskandarani, F. (2023). Inaccuracy of Machine Translation in Translating Medical Terms from English to Arabic. *Faculty of Languages Journal-Tripoli-Libya*, *1*(28), 234–210. https://doi.org/10.56592/flj.v1i28.818
3. Al-Jarf, R. S. (2018). Multiple Arabic Equivalents to English Medical Terms. *International Linguistics Research*, *1*(1), p102. https://doi.org/10.30560/ilr.v1n1p102
4. Alsunousi, Z. A. M. (2025). Investigating recurrent translation errors in English-to-Arabic translation by undergraduate students at the Faculty of Languages and Translation, Misurata University. *The Libyan Journal of Language, 6*(1), 1–9. https://azujournals.ly/index.php/LJ/article/view/79

5. Bowker, L. (2020). Chinese speakers' use of machine translation as an aid for scholarly writing in English: a review of the literature and a report on a pilot workshop on machine translation literacy. *Asia Pacific Translation and Intercultural Studies*, *7*(3), 288–298. https://doi.org/10.1080/23306343.2020.1805843

6. Bowker, L., & Ciro, J. B. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Emerald Publishing Limited. https://doi.org/10.1108/9781787567214

7. Cai, L., Tan, H., & Huang, M. (2025). Technology Competence Training of Translation Educators in the Age of Technology-and-Information Empowerment: A Chinese Perspective. *SAGE Open*, *15*(3). https://doi.org/10.1177/21582440251365384

8. Delfani, J., Orasan, C., Temizöz, Ö., Taylor-Stilgoe, E., Kanojia, D., Braun, S., & Schouten, B. (2024). *Google Translate Error Analysis for Mental Healthcare Information: Evaluating Accuracy, Comprehensibility, and Implications for Multilingual Healthcare Communication*.

9. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale Development: Theory and Applications*. SAGE Publications. https://books.google.com.ly/books?id=QddDEAAAQBAJ

10. European Commission. (2022). *European Master's in Translation: Competence framework 2022*. https://commission.europa.eu/document/download/b482a2c0-42df-4291-8bf8-923922ddc6e1_en?filename=emt_competence_fwk_2022_en.pdf

11. Hackenbuchner, J., & Krüger, R. (2023). DataLitMT – Teaching Data Literacy in the Context of Machine Translation Literacy. *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 285–293). European Association for Machine Translation. https://aclanthology.org/2023.eamt-1.28/

12. Hazaea, A. N., & Qassem, M. (2024). Translation Competence in Translator Training Programs at Saudi Universities: Empirical Study. *Open Education Studies*, *6*(1). https://doi.org/10.1515/edu-2024-0029

13. Huang, J. (2023). Exploring Students' Perception of Subject and Translation Integrated Learning (STIL) as an Innovative Approach to Translation Instruction: A Case Study. *Sage Open*, *13*(4). https://doi.org/10.1177/21582440231212261

14. Hurtado Albir, A. (2017). *Researching translation competence by PACTE Group*. John Benjamins Publishing Company. https://doi.org/10.1075/btl.127

15. Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

16. Kovács, G. (2023). Medical Texts and Their Translation in Translator Training. *Acta Universitatis Sapientiae, Philologica*, *15*(2), 75–85. https://doi.org/10.2478/ausp-2023-0017

17. Krüger, R. (2022). Integrating professional machine translation literacy and data literacy. *Lebende Sprachen*, *67*(2), 247–282. https://doi.org/10.1515/les-2022-1022

18. Krüger, R. (2023). Artificial intelligence literacy for the language industry – with particular emphasis on recent large language models such as GPT-4. *Lebende Sprachen*, *68*(2), 283–330. https://doi.org/10.1515/les-2023-0024

19. Krüger, R. (2024). Outline of an Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication. *Lublin Studies in Modern Languages and Literature*, *48*(3), 11–23. https://doi.org/10.17951/lsmll.2024.48.3.11-23

20. Mahdi, A. I., & Mohammed, H. G. (2025). Accuracy of Medical Terminologies Translated from English into Arabic: The Case of Audiovisual Shows. *International Journal of Arabic-English Studies*. https://doi.org/10.33806/ijaes.v25i2.506

21. Man, D., Zhu, C., Chau, M. H., & Maruthai, E. (2022). Contextualizing assessment feedback in translation education: A corpus-assisted ecological approach. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1057018

22. Mekheimer, M. (2026). A diagnostic assessment of translation competence: linguistic and intercultural gaps in Egyptian EFL programs. *Language Testing in Asia*, *16*(1), 9. https://doi.org/10.1186/s40468-025-00423-3

23. Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *iScience*, *27*(10), 110878. https://doi.org/10.1016/j.isci.2024.110878

24. Prieto Ramos, F., & Guzmán, D. (2024). The impact of specialised translator training and professional experience on legal translation quality assurance: an empirical study of revision performance. *The Interpreter and Translator Trainer*, *18*(2), 313–337. https://doi.org/10.1080/1750399X.2024.2344948

25. Rafah, A. A., & Elraggas, A. A. (2024). The difficulties of translating English phrasal verbs into Arabic: The case of Libyan EFL graduate translation students at the Libyan Academy. *Afaq Journal for Human and Applied Studies*, (2), 428–454.

26. Rao, P., McGee, L. M., & Seideman, C. A. (2024). A Comparative assessment of ChatGPT vs. Google Translate for the translation of patient instructions. *Journal of Medical Artificial Intelligence*, *7*, 11–11. https://doi.org/10.21037/jmai-24-24

27. Salamah, D. (2021). Translation Competence and Translator Training: A Review. *International Journal of Linguistics, Literature and Translation*, *4*(3), 276–291. https://doi.org/10.32996/ijllt.2021.4.3.29

28. Saleh, K. A. A. (2026). Challenges faced by Libyan translation students when translating scientific texts from English into Arabic. *The Libyan Journal of Language,* *7*(1), 68–82. https://azujournals.ly/index.php/LJ/article/view/180

29. Strasly, I. (2026). A project-based approach to teaching medical translation to deaf students: integrating the BabelDr corpus in a continuing education programme at the faculty of translation and interpreting, University of Geneva. *The Interpreter and Translator Trainer*, 1–21. https://doi.org/10.1080/1750399X.2026.2618880

30. Xie, H., & Wang, P. (2025). Metacognition and translation competence: A three-level meta-analytic study. *Acta Psychologica*, *260*, 105722. https://doi.org/10.1016/j.actpsy.2025.105722

31. 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

Appendices

Appendix A. Medical Translation Task

Student code: ..... Semester: Five / Six / Eight Time: 45 minutes

Instruction: Translate the following text into clear and accurate Arabic for adult patients and their families. Preserve every instruction, qualification, number, and negative statement. Do not add medical advice. Follow the resource-use condition announced by the researcher.

Taking Antibiotics Safely

Antibiotics are medicines used to treat infections caused by bacteria. They do not work against viral infections such as the common cold or influenza. If your doctor prescribes an antibiotic, take the exact dose at the recommended time and complete the prescribed course, even if your symptoms improve. Stopping treatment early may allow some bacteria to survive and can contribute to antimicrobial resistance. Tell your doctor if you have previously had an allergic reaction to an antibiotic. Seek urgent medical help if you develop difficulty breathing, swelling of the face or throat, or a severe skin rash. Mild effects such as nausea or diarrhea may occur, but persistent or bloody diarrhea should be reported promptly. Do not share antibiotics with another person and do not use medicine left from an earlier illness. Some antibiotics interact with other medicines, so inform the healthcare professional about all prescription medicines, non-prescription products, and supplements you use. If you miss a dose, follow the instructions on the label or ask a pharmacist; do not take a double dose unless specifically instructed.

Appendix B. Analytic Scoring Rubric

Criterion	Maximum	Full-credit performance
Meaning accuracy	30	All propositions, relations, instructions, negation, time, quantity, and severity are preserved without distortion.
Medical terminology	25	Terms are accurate, contextual, current, and consistent; abbreviations are correctly interpreted.
Arabic language quality	20	Grammar, syntax, spelling, punctuation, cohesion, and natural expression are controlled.
Style and audience appropriateness	15	Register is clear and suitable for adult patients without loss of technical meaning.
Completeness and consistency	10	No unjustified omission or addition; repeated concepts and formatting remain consistent.

Rater comments: _____

_____

_____

Appendix C. Supporting Questionnaire

Use the same anonymous code as the translation task. This questionnaire records background and translation practices.

1. Semester:

   Five / Six / Eight

2. Have you previously studied specialized or medical translation?

   Yes / No

3. How often do you use digital translation resources?

   Rarely / Sometimes / Often

4. Which resources do you normally use?

   Specialized dictionary / General dictionary / Terminology database / Search engine / Machine translation / Corpus / Other

5. Which difficulties did you experience?

   Multiple Arabic equivalents / Medical abbreviations / Limited medical knowledge / Complex English syntax / Evaluating source reliability / Producing natural Arabic / Maintaining consistency / Revising machine output