Construction and Empirical Study of an English Major Translation Competence Assessment Mechanism

Xiaoli Yu

English College, Zhejiang Yuexiu University, Shaoxing, Zhejiang, China. 312000

Abstract: English Major translation competence assessment has long relied on outcome-based evaluations such as traditional final exams, which suffer from single dimensionality, lack of process diagnosis, and strong subjectivity. Based on the MQM multidimensional quality model and generalizability theory, this study constructs a three-in-one translation competence assessment mechanism for English majors, covering linguistic, cultural, and strategic dimensions. A Chinese-English experiment was conducted with 60 second-year or third-year English majors, with 3 raters scoring independently, and generalizability theory was used to analyze error components. Results show a G coefficient of 0.82, exceeding the excellent threshold of 0.80. Person variance accounts for 48.7% of the total variance, much higher than rater variance and residual error. Major errors among English majors are concentrated in language conventions (34.1%) and style (28.7%). This study provides a scientific, operable, and diagnostic assessment tool for English major translation competence.

Keywords: translation competence assessment; MQM model; generalizability theory; English majors

DOI:10.12417/3029-2328.26.05.005

1.Introduction

For a long time, translation competence assessment for English majors has been outcome driven. It lacks process diagnosis an a multidimensional evaluation framework.

The situation right now is far from ideal. The mainstream method still boils down to final exam marks or teacher ratings, which are highly subjective. There is no standardized tool that covers multiple dimensions. Translation courses do exist for English majors, but the assessment methods are largely the same across the board. This makes it hard to pinpoint where a student is actually struggling, whether in language, culture, or strategy. Translation tests carry heavy subjectivity, and rater error alone eats up about 30% of total variance, which seriously undermines reliability. Because of the bias toward certain teaching approaches, translation competence has stayed at the margins of the English major curriculum, and students end up with a systematic lack of translation training [1].

The problem is not just about testing methods. It runs through everyday teaching practice too [2]. This study sets out to answer a core question. How do you build a translation competence assessment mechanism for English majors that is scientific, practical, and actually fits the learners? To get at this, a functionalist perspective on translation theory helps. In that framework, designing an assessment means paying attention to three things at once. Whether the original text's function comes through. Whether the target reader can accept the result. And how raters interact as an assessment community.

2.Theoretical Foundations

2.1 Conceptual Framework of Translation Competence

Translation competence is the core of translation assessment. The PACTE model decomposes it into bilingual competence, extra-linguistic competence, translation knowledge, operational competence, and strategic competence[3]. For English majors, China's Standards of Translation Competence require the general translation level, i.e., the ability to handle daily communication and general text translation, with basically accurate and fluent target texts and the ability to use basic translation strategies[4].

2.2 House's Register Equivalence Model

Juliane House's translation quality assessment model based on systemic functional linguistics judges register equivalence between the source and target texts from three dimensions: field, tenor, and mode[5]. This model is used

to determine what to assess, i.e., whether the translation is functionally equivalent. The core of House's model lies in judging the degree of functional realization by matching the field, tenor, and mode of the source and target texts, which provides an important theoretical reference for the dimensional division of this study's assessment mechanism[6].

2.3 The MQM Multidimensional Quality Model

The MQM model classifies translation problems into seven dimensions: terminology, accuracy, language conventions, style, regional conventions, audience adaptability, and design and markup, supporting error classification and weighted scoring. The functionalist orientation of the MQM model is consistent with Nida's functional equivalence theory[7]. Its fundamental criterion for assessment is the degree to which the target text achieves the intended communicative function in a specific context, rather than mere word-for-word correspondence. With its systematic error classification system and modular design, the MQM model enables multidimensional quantitative assessment of translation quality, and its precision has been widely recognized by both the translation industry and research community[8]. This study adopts its error classification framework for how to quantify assessment.

2.4 Generalizability Theory

Generalizability theory assesses reliability by decomposing measurement errors into multiple sources and calculating the G coefficient[9]. This study adopts a "person-task-rater" three-facet crossover design to ensure reliability. The three components form a complete logical chain: House's model determines assessment content, the MQM model quantifies the assessment method, and generalizability theory controls assessment reliability.

3.Construction of the Assessment Mechanism Model

3.1 The Three-in-One Assessment Model

This assessment mechanism comprises three dimensions: the linguistic dimension, weighted at 40%, assesses translation accuracy and normative quality, covering errors such as omissions, mistranslations, spelling mistakes, and grammatical errors; the cultural dimension, weighted at 30%, assesses the ability to handle cultural differences, covering culture-loaded term translation and audience adaptability. Whether in traditional text translation or emerging fields such as audiovisual translation, the assessment of translation accuracy has always been the core of quality assurance systems[10], which provides a theoretical basis for this study to take linguistic accuracy as the primary assessment dimension.The strategic dimension, weighted at 30%, assesses the translator's ability to use translation strategies, including two sub-dimensions: use of translation strategies and post-editing capability. Particularly when translating texts involving specialized disciplinary knowledge, language normative quality requires not only general grammatical correctness but also conformity to the professional expression conventions of that discipline, which is especially important in the translation assessment of English majors[11]. Use of translation strategies involves the appropriate choice between literal and free translation, as well as the proper application of domestication and foreignization. Post-editing capability involves the ability to modify and optimize machine-translated or initial draft texts. Table 1 presents the three dimensions of the proposed assessment mechanism, including their respective weights, sub-dimensions, and specific assessment content.

Table 1. Three dimensions of the assessment mechanism and their sub-dimensions

Dimension	Weight	Sub-dimension	Assessment Content
Linguistic dimension	40%	Accuracy	Omission, mistranslation, addition
Linguistic dimension	40%	Language normativity	Spelling errors, grammatical errors, improper word order
Cultural dimension	30%	Culture-loaded term translation	Accurate conversion of Chinese characteristic expressions

Cultural dimension	30%	Audience adaptability	Whether the translation conforms to the target readers' cultural cognition and language habits
Strategic dimension	30%	Use of translation strategies	Appropriate choice between literal and free translation, proper application of domestication and foreignization
Strategic dimension	30%	Post-editing capability	Modification and optimization of machine-translated or initial draft texts

(Note: Continued Table 1)

3.2 Scoring Method

This study adopts the MQM error classification weighted scoring method, dividing error severity into three levels: minor errors penalized 1 point (spelling errors, punctuation errors), major errors penalized 5 points (grammatical errors, improper word order), and critical errors penalized 10 points (omission of core information, serious terminology mistranslation). Total penalty points equal the sum of the number of errors of each type multiplied by their corresponding penalties. The translation error rate equals total penalty points divided by the number of characters in the translation multiplied by 100%. The translation quality score equals 100 minus the translation error rate. To control rater subjectivity, this study adopts an independent multi-rater scoring design and estimates rater variance components and calculates the generalizability coefficient G coefficient as a reliability indicator through generalizability theory. A G coefficient greater than 0.80 indicates excellent reliability[12].

3.3 Operational Procedure

The operational procedure of this assessment mechanism consists of four steps. The first step is translation task implementation, in which students complete a paragraph translation from Chinese to English within 30 minutes, with the text focusing on Chinese culture, 140 to 160 characters in length, and difficulty level equivalent to that of a standard paragraph translation task of moderate difficulty.The second step is MQM error annotation, in which raters, after receiving training in MQM error classification, annotate each translation sentence by sentence, recording the error type, location, and severity level. The third step is multi-rater scoring and generalizability analysis, in which at least 2 to 3 raters score independently, adopting a person-task-rater three-facet crossover design, and using GENOVA software to estimate variance components and calculate the G coefficient. The fourth step is comprehensive assessment report output, in which the system generates a diagnostic report including total score, dimension-specific scores, error type distribution chart, and capability profile, providing targeted feedback to students.

4.Experiment and Analysis

4.1 Experimental Design

This study takes 60 second-year or third-year English majors (including Translation, Business English, and English majors) who have completed basic English courses and Chinese-English translation courses as subjects, with an average age of 19.5 years. The students come from three programs: Translation, Business English, and English. The translation task is a 150-character Chinese-to-English paragraph on chopstick culture, of moderate difficulty and containing culture-loaded terms. Sixty translations were collected and each was scored independently by three raters, giving 180 data points in total. MQM error annotation and penalty statistics were also carried out.

4.2 Error Distribution

Across the 60 translations, a total of 1,247 errors were annotated, with an average of 20.8 errors per translation. Table 2 presents the full distribution by error type and dimension.

Table 2. Distribution of student translation errors by dimension

Dimension	Error Type	Error Count	Percent-age	Sub total
Language conventions	Spelling errors	156	12.5%	34.1%

Language conventions	Grammatical errors	189	15.2%	34.1%
Language conventions	Improper word order	79	6.3%	34.1%
Language conventions	Punctuation errors	49	3.9%	34.1%
Style	Unnatural expression	248	19.9%	28.7%
Style	Organizational style	113	9.1%	28.7%
Accuracy	Omission	156	12.5%	20.4%
Accuracy	Mistranslation	98	7.9%	20.4%
Culture	Audience adaptability	143	11.5%	11.5%
Terminology	Terminology mistranslation	42	3.4%	3.4%
Total		1,247	100%	100%

(Note: Continued Table 2)

In the language conventions dimension, spelling errors accounted for 12.5% (156 errors), grammatical errors 15.2% (189 errors), and improper word order 6.3% (79 errors). This dimension alone made up 34.1% of all errors, which lines up with what other studies have found about English majors. Their main translation weakness sits at the basic level of vocabulary and grammar, a sign that foundational language skills are not yet solid [13].

In the style dimension, unnatural expression accounted for 19.9% (248 errors) and organizational style for 9.1% (113 errors), together making up 28.7% of the total. Unnatural expression errors come mainly from three sources, namely native language thinking, too much reliance on word for word matching, and difficulty switching between Chinese and English syntactic structures and expression habits [14]. Accuracy errors, which cover omission at 12.5% (156 errors) and mistranslation at 7.9% (98 errors), totaled 20.4%. Audience adaptability errors in the cultural dimension accounted for 11.5% (143 errors), while terminology mistranslations made up 3.4% (42 errors).

A single translation example shows how multiple error types can pile up in one sentence. The source text says that chopsticks are uniquely distinctive tableware of the Chinese nation, possessing unique features compared to tableware from other countries. One student wrote it as "Chopsticks are Chinese nation unique tableware, compared with other countries' tableware has unique characteristics." In just this one sentence, three things go wrong, a spelling mistake, an incorrect noun premodification, and a redundant syntactic structure.

A similar pattern comes up in cultural errors. The phrase "a country of etiquette" was often translated literally, with students missing its cultural connotations entirely. This suggests that audience adaptability errors in the cultural dimension go beyond conceptual understanding. They also have to do with how target language readers process and receive culturally loaded information.

4.3 Generalizability Theory Analysis

Generalizability theory can quantify and decompose error sources such as person ability differences, rater stringency differences, and task difficulty differences in translation scoring, and its decision studies can provide precise basis for optimizing the number of raters and tasks. A three-facet crossover design of person, task, and rater is adopted, and GENOVA software is used to estimate variance components. The results show that person variance accounts for 48.7% of the total variance, indicating that the assessment mechanism effectively reflects real differences in students' translation competence. Rater variance accounts for 17.9% of the total variance, indicating that the rater effect is significant and a multi-rater design is necessary. Residual error accounts for 24.5% of the total variance, and task variance accounts for only 8.5%, indicating that the translation task has good representativeness.

The generalizability coefficient G coefficient is calculated as a reliability indicator. Under the current design of one task and three raters, the G coefficient is 0.82, reaching the excellent level[15]. Decision studies show that with one rater, the G coefficient is 0.64, indicating insufficient reliability. With two raters, the G coefficient is 0.76, which

is acceptable. With three raters, the G coefficient is 0.82, indicating excellent reliability. Increasing the number of raters beyond this point yields limited reliability improvement, with the G coefficient reaching at most 0.87. Therefore, it is recommended to use at least 2 to 3 raters.

Figure 1 presents the G coefficients with different numbers of raters. The horizontal dashed line indicates the excellent reliability threshold of 0.80. The figure shows that the G coefficient reaches 0.82 with three raters, after which further increases in the number of raters yield only marginal improvement.

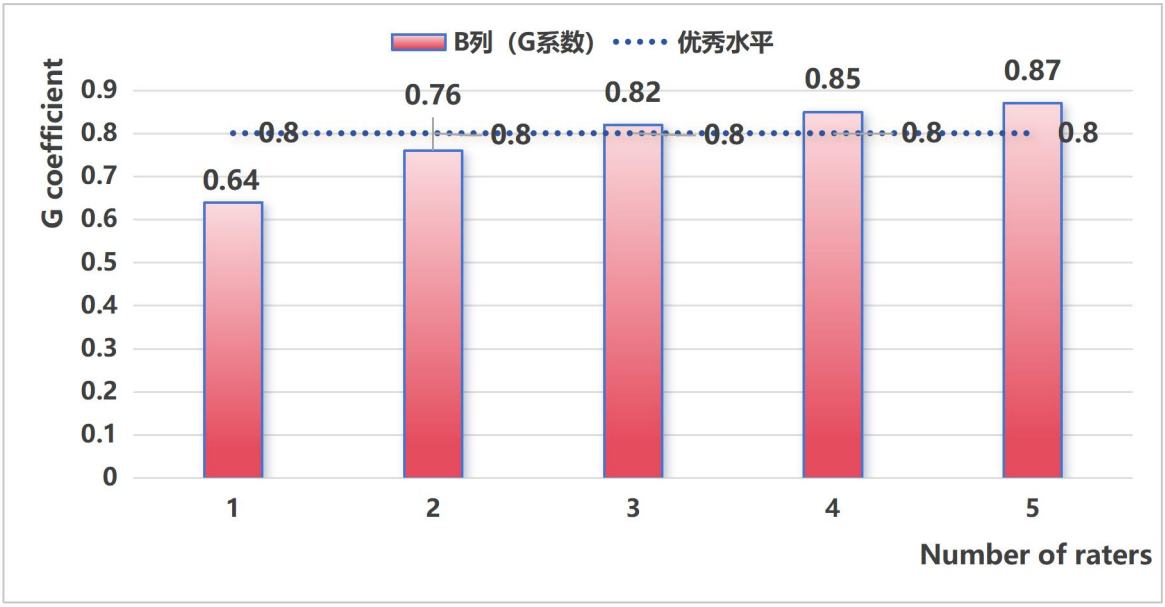

Figure 1. G coefficients with different numbers of raters

5.Results Evaluation

5.1 Validity Analysis

This assessment mechanism covers three dimensions of language, culture, and strategy, encompassing the core requirements of the general translation level in China's Standards of Translation Competence[16], and thus has good content validity. The competence descriptions in the English major translation syllabus correspond precisely to the external reference standard for the content validity of the three dimensions—language, culture, and strategy—in this study's assessment mechanism.

The generalizability analysis tells a clear story. Person variance accounts for 48.7% of total variance, well above the rater variance of 17.9% and residual error of 24.5%. In plain terms, the scores are picking up real differences in translation competence, not just measurement noise. That is a good construct validity result.

5.2 Comparison with Traditional Assessment Methods

Compared with single-score assessment, this mechanism has three advantages. The most important one is diagnostic capability. It provides a total score but also gives dimension specific scores and error type distributions, so students can actually see where their weaknesses lie. Second, objectivity improves because the multi rater design and generalizability theory framework bring scoring error under control. The G coefficient rises from 0.64 to 0.82, a meaningful reliability gain [17]. Third, the procedure is standardized with clear criteria, so it can work in real teaching settings, not just in research.

5.3 Optimization Suggestions

Based on the decision study results, I would suggest four things.

One recommendation is to regularly check learners' emotional states, such as foreign language learning burnout, and adjust task difficulty or feedback accordingly when designing assessments. There is empirical evidence that

emotional states can significantly moderate how much translation competence improves under different teaching modes [18].

For rater configuration, at least two to three raters should be used, and three is preferred when conditions allow. For task configuration, two translation tasks on different themes, such as cultural and technological topics, can further lift reliability. Prior work has identified task design and rater training as the two critical factors in translation assessment reliability [19].

On the intelligent assistance side, I would suggest drawing on existing translation competence standards to build clear, multi level evaluation criteria and reduce the subjectivity of manual scoring. The comparative work on China's Standards of Translation Competence and the translation competence scale from China's Standards of English offers a useful reference for aligning indicators with national benchmarks [20].

In the long term, a dynamic database tracking students' translation competence development should be built, fitting individual growth trajectories through data models. This would shift assessment from a one time grade to ongoing competence tracking.

6.Conclusion

This study constructs and empirically tests a three in one translation competence assessment mechanism for English majors. The mechanism has high reliability and validity. Its G coefficient reaches 0.82 and person variance accounts for 48.7% of total variance, so it can effectively diagnose students' translation weaknesses.

The main weaknesses concentrate in two dimensions. Language conventions account for 34.1% and style accounts for 28.7%. This points clearly to where teaching should focus. Strengthening translation theory, updating teaching methods, and reinforcing language competence cultivation all have practical value here. A more concrete measure is to increase comparative practice on literal and free translation, helping students develop scientific decision making between faithfulness and smoothness.

The rater effect is a major error source, with rater variance at 17.9% of total variance. At least two to three raters are needed to ensure reliability, and this is not optional.

This study offers translation teaching a tool that is scientific, operable, and diagnostic. It lays a methodological foundation for improving English majors' translation competence and international communication literacy. Future research could combine generative AI tools such as DeepSeek and ChatGPT for preliminary translation evaluation, followed by human rater review and editing, striking a balance between efficiency and quality. Under ecological translation theory, the translator acts as an ecological regulator, attending not only to linguistic conversion but also to cultural recognition and discourse adaptation, so the translation achieves effective embedding and ecological balance in the cultural context.

References:

[1] Li Z,Kong D H,Deng G S.Research on English majors'translation competence for international communication: Empirical analysis and improvement paths[J].Journal of Anhui University of Science and Technology(Social Science Edition),2026,28(2):71-78.

[2] Zhang Y J.A study on the cultivation of students'translation competence in current English major teaching[C]//New Curriculum Research Magazine.Proceedings of the High-Quality Education Development Conference Focusing on New Curriculum Reform(Vol.2).New Curriculum Research Magazine,2023:37-38.

[3] Wang C.English-Chinese translation of Evaluating the Evaluator:A Novel Perspective on Translation Quality Assessment(Chapter 2)and a report on the translation[D].Guizhou Normal University,2022.

[4] Xia H.A report on the E-C translation of English Language Learning and Technology(excerpt)based on Juliane House's revised model of translation quality assessment[D].Liaocheng University,2023.

[5] Feng X R.An analysis of translation methods for long and difficult sentences in English major under the functional equivalence theory[J].English Square,2023(25):32-35.

[6] Zhou C L.A study on Portuguese-Chinese translation quality assessment and post-editing based on generative AI and NMT[J].Science and Technology Communication,2026,18(7):35-41.

[7] Tong Q Q,Cheng J S.A review of quantitative assessment models for audiovisual subtitle translation quality abroad[J].Journal of Ankang University,2025,37(6):60-67.

[8] Du D S.Translation quality evaluation of English literature for science and technology for English majors[J].Technology Innovation and Productivity,2021(11):26-28+31.

[9] Cui L.Existing problems and reform measures in the translation competence of English major normal students:A case study of Yan'an University[J].Shaanxi Education(Higher Education),2025(9):33-35.

[10] Guang Y H.The current state and factor analysis of Chinese-English written translation ability of university English learners[J].Journal of Nanyang Institute of Technology,2025,17(3):49-54.

[11] Xu Z R,Wang L L.A study on the improvement paths of college students'English translation competence from the perspective of international communication[J].English Square,2024(26):107-111.

[12] Yan L,Zhang W,Xiang G H.An application study of generalizability theory in English major translation scoring[J].Journal of Hubei University of Economics(Humanities and Social Sciences Edition),2021,18(11):93-96.

[13] Dai Z L.English major translation teaching based on China's Standards of Translation Competence[J].Shanghai Journal of Translators,2022(5):54-57.

[14] Liu Q,Li Q P.A study of the relationship among different online teaching modes,foreign language learning burnout,and English major learners'translation competence[J].Foreign Languages and Translation,2025,32(3):81-88.

[15] Li R.An intelligent evaluation algorithm for the translation quality of Chinese-to-English paragraphs[J].Journal of Hunan Industry Polytechnic,2023,23(4):121-125.

[16] Liu Y F.Evaluation of machine translation tools and optimization of teaching strategies in English major translation teaching:A case study of ChatGPT-4o[J].Modern English,2024(24):121-123.

[17] Sun F J.Theory and practice of English major translation testing in the information technology environment[J].Food Research and Development,2022,43(2):235.

[18] Zheng J S.A study on strategies for improving college students'English translation competence[J].Overseas English,2024(5):138-140.

[19] Liu X J.Generative AI-empowered literary translation quality assessment and strategy optimization:An analysis of ecological transformation in DeepSeek's translation of Faulkner's The Town[J].Modern Chinese Literature and Culture,2026(6):108-110.

[20] Mu L,Liang W L,Liu X Y.A comparative study of China's Standards of Translation Competence and the translation competence scale of China's Standards of English[J].Chinese Foreign Language,2024,21(3):87-97.