

# Quantitative Assessment of Translation Quality in Education, Certification, and Industry: An Overview

Jing Wang<sup>1</sup>, Qiushi Gu<sup>2\*</sup>

<sup>1</sup>Artificial Intelligence and Language Cognition Laboratory, Beijing International Studies University, Beijing, China

<sup>2</sup>Research Office, Beijing International Studies University, Beijing, China

Email: \*guqiushi@bisu.edu.cn

**How to cite this paper:** Wang, J., & Gu, Q. (2024). Quantitative Assessment of Translation Quality in Education, Certification, and Industry: An Overview. *Open Journal of Modern Linguistics*, 14, 209-223. <https://doi.org/10.4236/ojml.2024.142012>

**Received:** March 11, 2024

**Accepted:** April 25, 2024

**Published:** April 28, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In an era of increasing globalization and multilingual interactions, translation quality assessment has become crucial. Yet, a notable gap persists in contemporary research, especially regarding standardization and systematic exploration to the quantitative assessment of translation quality, an area that has not garnered significant scholarly focus. This research provides a thorough examination of the methods and standards utilized for assessing translation quality across educational, certification, and industrial domains. Emphasizing the pivotal role of translation quality assessment in ensuring the accuracy, fluency, coherence, and appropriateness of translated materials, this paper sheds light on the existing disparities in scoring methodologies prevalent within educational and certification settings. Furthermore, it also delves into the assessment frameworks employed in industry, revealing their complexity and the detailed evaluation criteria for both human and machine-translated texts. To improve teaching efficiency and translation quality, the study proposes integrating academic and industry assessment criteria. Moreover, it envisions the evolution of future translation quality assessment models, envisaging the incorporation of advanced functionalities such as automated error detection and instantaneous feedback mechanisms aimed at optimizing the translation output.

## Keywords

Translation Quality, Quantitative Assessment, Educational Testing, Certification Exam, Industry Standards

## 1. Introduction: Theoretical Background and Practical Significance of Translation Quality Assessment

Translation quality assessment (TQA) entails a systematic and thorough appraisal of translations, grounded in predefined evaluative standards, and constitutes a pivotal area of research within the field of translation studies. This area of inquiry has garnered considerable attention since the late 20<sup>th</sup> century, notably influenced by scholars including Brunette (2000), Colina (2008), House (1981, 2002), Maneesriwongul and Dixon (2004), Si (2004, 2008), Wu (2007), Wang and Zhu (2016). TQA establishes a comprehensive approach, leveraging both systematic, subjective, and objective parameters to assess translation quality, accuracy, delivery, and the exact transfer of the source text's essence.

The paradigm of TQA transcends mere linguistic translation, undertaking an extensive analysis involving cultural nuances, contextual relevance, and specific intent behind translations. Typically, TQA is categorized into qualitative and quantitative approaches (Yang, 2012). Qualitative TQA explores subjective aspects such as stylistic consistency and cultural relevance, using interpretative analysis and expert insights for evaluating the effectiveness of translations in conveying the original narrative. This approach considers clarity, cultural adaptability, and target audience relevance. Quantitative TQA, on the other hand, employs measurable, objective criteria, utilizing models like parameter-based, error deduction, and competence description frameworks. This facet provides a structured, bidirectional evaluation, supported by statistical tools, readability indices, and adherence to established norms, offering a replicable framework for assessing translation quality (Yang, 2012). While the former involves assessing translations against certain models without yielding numerical values, the latter emphasizes a quantifiable presentation of outcomes, distinguishing the merits and limitations of translations. The quantitative analysis of translation quality subdivides into assessments of human and machine-generated translations, with the former employing manual scoring under prescribed quality standards and the latter employing algorithmic methodologies for automated grading. Studies on traditional TQA began earlier, resulting in a relatively mature theoretical and practical framework. These manual assessment systems are primarily applied in educational testing, national and international translation proficiency assessments, such as China Accreditation Test for Translators and Interpreters (CATTI), The Canadian Language Quality Measurement System (Sical), and the American Translators Association (ATA) Certification Exam.

Exploring quantitative evaluation of translation quality holds paramount importance within the realms of translation education, certification processes, and industry. Presently, various institutions and scholars employ diverse methods for quantitative assessment in those settings. This paper endeavors to conduct an in-depth exploration and analysis of these varied practices, focusing on their characteristics, operational methodologies, and the potential for synergistic ef-

fects. The primary objective is to analyze representative quantitative methods used in educational settings, certification examinations, and industrial contexts. This comprehensive examination aims to understand the strengths and limitations of each model thoroughly and provide pragmatic recommendations for their application. Such an investigation is crucial not only for enhancing the efficacy of translation education and certification but also for advancing the overall quality standards within the translation industry. By examining these practices from a global perspective, the study seeks to contribute to the ongoing discourse on translation quality assessment, highlighting the interplay between various assessment techniques and their cumulative impact on the translation field.

## 2. TQA in Education<sup>1</sup>

Focusing on TQA in education, this section selects the scoring standards of the translation part in China's College English Test Band Four and Six (CET-4, CET-6) and Test for English Majors Band-Eight (TEM-8) as the subjects of analysis. As large-scale, standardized tests encompassing a broad student demographic, these examinations present ideal objects of study. The choice to analyze these specific tests is grounded in their widespread recognition and their established role in shaping translation competency standards, reflecting prevalent trends in educational translation assessment.

According to *Syllabus for College English Test—Band Four and Six (CET-4, CET-6) (Revised in 2016)* (National Education Examination Authority, n.d.), the translation section primarily involves Chinese-to-English translation tasks covering various themes. It constitutes 15% of the total score, segmented into five scoring tiers: 14, 11, 8, 5, and 2 points. These criteria, while detailed for each tier, largely assess overall translation quality—emphasizing fidelity, fluency, structural clarity, and language accuracy. Examiners, trained comprehensively, utilize standardized exemplars for scoring in order to ensure uniformity.

The TEM-8 exam adopts a more refined grading standard, categorizing translations into levels ranging from excellent, good, satisfactory, unsatisfactory, and very poor. Scoring involves considering the fidelity to the source text, which contributes to 70% of the assessment, and linguistic appropriateness, accounting for the remaining 30%. This standard emphasizes the fidelity of the translation to the source text, word choice, syntax, spelling, and punctuation, and the readability of the translation (Chen, 2016). Additionally, the test provides the conventional deducting-mark scoring approach for mistranslations and errors based on specific suggested answers (Zhao & Cai, 2018). Although these standards reflect the multidimensional quality of translations to a certain extent, they primarily rely on overall scoring, lacking quantified details for

<sup>1</sup>In this paper, “TQA in translation education” refers to the inclusion and assessment of translation skills in broader educational domains, such as in university-level language exams like CET-4, CET-6, and TEM-8.

scoring.

From the analysis above, it is evident that China's translation education scoring standards predominantly use an overall scoring method, lacking detailed quantification of scoring aspects. While this approach, to a certain extent, reflects the complexity and multidimensionality of translation scoring, it places significant demands on the examiners' capabilities, emphasizing the need for objectivity to ensure consistent and reliable scoring. Simultaneously, this method reveals deficiencies in objectivity and repeatability, lacking clear guidance for scoring details. Consequently, this scoring model can be perceived as a relatively rudimentary reference within the scoring paradigm (Yang, 2012). Therefore, to improve objectivity and consistency in scoring, there is an exigent need in the realm of the educational TQA to integrate more sophisticated, detailed, and quantified scoring standards.

### 3. TQA in Certification Exams

TQA in qualification examinations constitutes a fundamental aspect of evaluating the proficiency of translators and the standard of their translation output. Distinct standards for assessing translation quality have been established by various countries and institutions, each focusing on specific elements of translation quality. In this part, emphasis is placed on specific standards such as error deduction systems due to their growing prominence in contemporary translation assessment practices. This section aims to offer an in-depth and detailed comparative analysis of the standards applied in prominent assessments of translation qualification, including CATTI (China), ATA Certification Exam (the U.S.), Chartered Institute of Linguists (CIOL) (UK) and Institute of Translation and Interpreting (ITI) (UK), Canadian Translators, Terminologists and Interpreters Council (CTTIC) (Canada), and National Accreditation Authority for Translators and Interpreters (NAATI) (Australia), revealing the commonalities and disparities in these diverse professional translation examinations (see **Table 1** for more details) and the integral role these diverse standards play in upholding translation quality globally.

CATTI (*China Accreditation Test for Translators and Interpreters, n.d.*), while not publicly disclosing its detailed grading criteria for the written test, clearly delineates the basic requirements for each level in its syllabus. These include adherence to the original text, linguistic adequacy, precision in terminology usage, textual fluency, and grammatical accuracy. This focus on content fidelity and linguistic normativity is echoed in recent studies emphasizing the necessity of such standards in maintaining translation integrity (Pym, 2010). The evaluation principles in CATTI align with those in CET-4, CET-6, and TEM-8, emphasizing content fidelity and linguistic normativity. Yet, these assessments lack granular parameters for detailed evaluation, inherently carrying a degree of subjectivity, a common trait in large-scale examinations such as CATTI.

**Table 1.** Table for TQA in qualification examinations.

Country	Name of the Exam	Assessment method	Key Features	Primary Parameters	Secondary Parameters
China	CATTI	Unspecified details	Focuses on fidelity, linguistic normativity	adherence to the original text, linguistic adequacy, precision in terminology usage, textual fluency, and grammatical accuracy	Unspecified details
The U.S.	ATA	Error deduction	Detailed error classification; emphasizes overall quality	Target language mechanics; meaning transfer; errors at word, phrase, sentence, and passage level, writing ability	Grammar, syntax, word form/part of speech, spelling, capitalization...; terminology, addition and omission, ambiguity, cohesion, faithfulness...; usage, text type, register, style...
UK	CIOL (DipTrans)	Level-based scoring system	Basic parameters included, suitable for large-scale tests	accurate transfer of information; grammatically sound, cohesive and coherent in the target language;	Comprehension, accuracy, transfer of names, dates and figures; Grammar, register style, linkages, discourse, orthography, punctuation, accentuation
UK	ITI	Error deduction with bonus points	Diverse scoring criteria, encourages excellent expressions.	Accurate transfer of content; target language form; excellent renderings	Terminology, register, collocation, grammar, syntax, rewording, spelling, punctuation, layout, presentation, omissions and additions, consistency, tense usage and tautology...
Canada	CTTIC	Error deduction method	Distinguishes between translation and language errors	Translation comprehension; language expression	Nonsense, omission, mistranslation, addition...; gibberish, non-idiomatic, syntax, ambiguity, repetition, spelling, punctuation...
Australia	NAATI	Level-based scoring system	Comprehensive assessment of translation and language competencies	Transfer competency (meaning transfer, application of textual norms and conventions); language competency	Consistency, distortions, omissions, addition; register, style, text structure, terminology; unidiomatic, grammar, syntax, spelling and/or punctuation...

In contrast, ATA employs a more detailed error deduction system for its certification exams. Examiners, who are required to be native speakers of the target language, follow an exhaustive error deduction flowchart and standardized error categorization framework. The ATA approach intricately classifies translation errors, defining a passing standard as a text with errors ranging from 0 to 17, while anything beyond is deemed unsatisfactory. The ATA's (*American Trans-*

lators Association, n.d.) assessment of completed translations takes into account both content and linguistic form against a detailed error deduction reference chart. The assessment of content encompasses a comprehensive examination of the target language's structural elements, encompassing errors in meaning transfer across individual words, sentences, and entire passages. This includes evaluating writing proficiency and factors such as fidelity to the source text, accuracy, coherence, redundancy, and omissions. The evaluation method precisely aligns verb tenses, terminology, word selection, register, and genre with the original text. Evaluations of linguistic form emphasize conformity to the syntactic and grammatical standards of the target language, inclusive of morphological structure, spelling accuracy, and punctuation consistency. Despite its foundation in error deduction, the ATA model primarily accentuates the overall quality of translations, balancing content faithfulness and accuracy with linguistic conformity. Such a comprehensive approach is increasingly recognized as crucial in ensuring translation accuracy and coherence.

The assessment criteria of the CIOL (*Chartered Institute of Linguistics, n.d.*), exhibiting greater detail compared to broader assessments like CATTI, provide practical and specific measures in translation evaluation. CIOL conducts a comprehensive assessment of translation skills in its Diploma in Translation (Dip-Trans) exam, catering to both general and specialized texts. The exam categorizes translations into four grades: Distinction (40 - 50 points), Merit (35 - 39 points), Pass (30 - 34 points), and Fail (0 - 29 points). The assessment standards specifically cover translators' comprehension, accuracy, and transfer of names, dates, and figures (50 points) as well as grammatical usage (50 points), including basic but critical parameters such as terminology, register, style, spelling, and punctuation.

Successful completion of the translation qualification examination administered by the ITI (*Institute of Translation and Interpreting, n.d.*) in the UK is a mandatory requirement for acquiring Qualified Translator Membership. The ITI exam features a multi-dimensional assessment system, balancing error deduction with an additional reward mechanism for exceptional translations. This dual approach is reflective of the evolving standard of translation quality, where creative expression is valued alongside technical accuracy. The grading system is divided into seven categories from A to G, with the first six being deductions and category G for bonus points. Examiners start deducting from a total of 95 points based on the assessment of each category, with a maximum deduction limit set for each. Following the deduction, examiners may award additional points for exemplary translations, with a passing score set at 63 points. ITI's evaluation encompasses a range of parameters like accuracy, terminology, register, collocation, punctuation, grammar, spelling, and omissions and additions, each described and exemplified for precision, ensuring the rigor of the grading framework. Furthermore, the guidelines stipulate that raters participate in comprehensive training programs to uphold objectivity and uniformity in evaluation, thereby enhancing the system's impartiality and equity.

CTTIC (*Canadian Translators, Terminologists and Interpreters Council, n.d.*)

adopts an error deduction approach, categorizing errors into translation and language errors, with each type penalized based on severity. Translation errors are divided into major (10 points deduction) and minor mistakes (5 points), while language errors are categorized into 3 levels, with major errors deducting 10 points, minor errors 5 points. Also, in the assessment of minor errors, deductions of 3 points are made for inaccuracies in spelling, punctuation, and typographical conventions. The cumulative score is determined through a weighted computation of these respective categories.

NAATI (*National Accreditation Authority for Translators and Interpreters, n.d.*) applies a five-band rating scale for its exam, assessing examinees' translation and language competencies. The evaluation focuses on skills in transfer competency, following meaning transfer skill, translation brief, and application of textual norms and conventions. It scrutinizes the translation's accuracy in conveying the source text's intent and content, its genre, style, and structural alignment with the source, alongside mastery of target language grammar, vocabulary, and syntax. From the perspective of language competency, such as the accuracy of the translation in conveying the source text's content and intent, along with the degree of consistency, omissions, additions, and distortions, are examined carefully. The genre, register, style, and structure of the translation are expected to match the source text. The assessment of language competency centers on the examinee's proficiency in the target language's grammatical structure, syntactic arrangement, lexical choice, orthographic accuracy, punctuation, and terminological precision, all of which are pivotal for effective meaning transfer.

From the analysis above, it is evident these standards ensure not just accurate information transfer from the source text but also the correct application of language, including terminology and grammar. While China's CATTI exam does not provide detailed public scoring criteria, other countries like the U.S., UK, Canada, and Australia offer comprehensive grading guidelines and criteria, facilitating in a more precise quality measurement of translations.

In particular, ATA, ITI, and CTTIC primarily implement error deduction systems, focusing on error identification in translations and deducting points based on error severity. This approach quantifies the assessment process and enhances grading objectivity. CIOL and NAATI utilize a reference based scoring system, grading translations against quality benchmarks or standard answers. It is noteworthy that ITI's exam, while deducting for errors, also rewards exceptional translation expressions, fostering translators' skill and creativity display. Additionally, NAATI's assessment extends beyond textual quality to evaluate the translator's overall competencies, underscoring the importance of abilities over mere literal accuracy.

Although ATA and ITI offer exhaustive error categorization and grading guidelines, with ATA even specifying a detailed grading process, such complexity and thoroughness in grading criteria could be demanding for examiners, especially when handling a large volume of translations within a short time. Since such methods may negatively impact the efficiency and practicality. Hence,



CIOL and CTTIC have adopted more simplified grading parameters, facilitating easier and more manageable application while maintaining quality assurance and efficiency in the evaluation process.

In sum, despite variances in assessment standards and methodologies across global translation qualification exams, their collective aim remains to ensure translation quality, continually exploring a balance among precision, fairness, and practicality in evaluations. As the translation field evolves with diverse demands, future grading standards are expected to incorporate varied assessment techniques, adapting to a broad spectrum and the changing landscape of translation tasks and outputs, including machine translation.

#### 4. TQA in the Industry

In the translation industry, quality assessment is a complex, multi-dimensional process aimed at ensuring the accuracy, fluency, and cultural appropriateness of translated works. The industry's focus on these key elements of TQA is underscored by their vital importance in upholding the integrity and efficacy of global communication. This attention reflects the shifting requirements and benchmarks within the translation sector, where high-caliber translations are essential for facilitating international business and cultural interactions. The standards for TQA vary by region, industry, and specific project requirements. Despite these variations, there are universally recognized benchmarks and guidelines (see **Table 2** for more details).

**Table 2.** Table for TQA in the industry.

Institution	Name	Applicable Field	Key Features	Relevant Parameters
The Standardization Administration of China (SAC)	GB/T-19682-2005	Translation service	Error rate formula, flexible weighting rates.	Fidelity to the source text, terminology consistency, fluency; proper nouns, measurement units, symbols, abbreviations, translation arrangement, semantic errors, omitted segments, mistranslations...
European Committee for Standardization (CEN)	EN15038	Certified translation services, specifically for European nations	Specific requirements for translators.	Terminology, grammar (including syntax, spelling, punctuation, orthography, diacritics), vocabulary, style, localization...
The Translation Bureau (Canada)	SICAL	Supports official large-scale translations	Error deduction; error categorization	Translation, language, terminology, (Unspecified details)
Localization Industry Standard Association (LISA)	LISA QA	Applicable to broader field of translation, localization.	Error deduction; error categorization	Accuracy, mistranslation, terminology, style, consistency, layout, font, images, figure numbering, descriptions, indexing, etc.
Society of Automotive Engineers (SAE)	SAE J 2450	Automotive industry, medical, industrial, and other terminology-intensive field; also applicable to machine translation	Error deduction; error categorization	Terminology, semantics, omissions, structural, spelling, punctuation, and other errors...



## Continued

The Translation Automation User Society (TAUS)	Dynamic Quality Framework (DQF)	Evaluates translation quality, applicable to localization industry management, suitable for machine translation	Error deduction; dynamic adjustment	Accuracy, fluency, compliance; accuracy rate, linguistic form, terminology, genre...
German Research Center for Artificial Intelligence (DFKI)	Multidimensional Quality Metrics (MQM)	Widely applicable, suitable for machine translation	Error deduction, error categorization, error rate formula, flexible weighting rates, simplified mode.	Accuracy, fluency, correct terminology, regional customs, style, content authenticity, format, and localization...

Language service industries have progressively established a series of standards to effectively guide translation activities worldwide. For instance, China issued the *Target Text Quality Requirements for Translation Service* (GB/T-19682-2005) in 2005 (National Public Service Platform for Standards Information, n.d.), and Europe introduced the *European Quality Standard for Translation Service Providers* (EN15038) in 2006 (BS EN-15038 European Quality Standard, n.d.). These frameworks have been instrumental in setting basic norms for the industry, specifying broad-based requirements for translators and establishing benchmarks for translation quality.

The GB/T-19682-2005, proposed by The Standardization Administration of China, meticulously lists relevant terms and definitions, prescribing essential criteria such as fidelity, terminology consistency, and fluency. It details specific requirements, including proper handling of proper nouns, units of measurement, symbols, abbreviations, and translation formatting. Predicated on the translation's intended purpose, the standard takes into account various related parameters as criteria for quality assessment, demanding an overall error tolerance rate not exceeding 1.5%. The classification of translation errors is methodically divided into 4 distinct categories, 1) Core semantic misunderstandings or misrepresentations, including key words (numbers), sentence or segment omissions, and mistranslations. 2) Semantic errors, including non-key words (numbers), omissions, mistranslations, and ambiguities in word choice or grammar. 3) Inaccuracies or inconsistencies in terminology, or mistranslations of proper nouns. 4) Units, symbols, and acronyms not translated according to established conventions. Relevant factors such as the intended use of the translation, the style and quality of the source text, professional complexity, and the length of time are considered in the assessment. The error rate formula provided in the standard varies according to these diverse variables such as different usage purposes, difficulties, and error types. This focus on terminological precision reflects the industry's increasing emphasis on uniform and accurate terminology usage, especially in specialized fields.

The EN15038 standard (BS EN-15038 European Quality Standard, n.d.), established by the European Committee for Standardization (CEN), is specifically designed for certified translation services within the European context, offering a commitment to quality in EU translation services. This standard also outlines a

comprehensive list of terms and definitions, encompassing requirements for the entire translation process, including various aspects of information involved in translation, roles of translators and reviewers, and the complete process from project initiation to final confirmation. Within its translation section, it outlines A to G, seven fundamental prerequisites, for translators, encompassing aspects such as terminology, grammatical accuracy (spanning syntax, spelling, punctuation, typography, diacritics), vocabulary, stylistic considerations, localization, formatting, and the overarching translation objective.

The Canadian Language Quality Measurement System (SICAL-Canadian Language Quality Measurement System) (Government of Canada, n.d.) is developed by the Translation Bureau of the Government of Canada. It forms the basis of the official TQA model, supporting large-scale translations and holding significant reputation and authority in Canada. SICAL categorizes translation quality into three acceptable levels, A. high quality, B. acceptable, and C. modifiable, and one unacceptable, D. unacceptable. This comprehensive system highlights the industry's need for clear, measurable standards in translation quality. It delivers descriptive definitions for each level across key facets like translation, language, and terminology, and stipulates an error range for each level, thereby injecting scientific precision and standardization into the assessment realm. For example, for a translation of 400 words from the original text, the absence of major errors and no more than six minor errors qualifies for level A, no major errors and up to twelve minor errors for level B, one major error and up to eighteen minor errors classifies it as level C, more than one major error or over eighteen minor errors for level D. The above scoring criteria serve as a reference for rating experts; however, in practical application, experts assess translation quality based on specific circumstances. This granularity in error classification and the emphasis on minor errors underline the increasing focus on detailed, nuanced assessment approaches in translation.

These national and international standard categories are comprehensive and rigorous, serving as universal benchmarks to evaluate translation quality across various domains and types. The broad adoption of TQA across diverse sectors signifies a growing global agreement on the need for organized and uniform methods in the translation industry. This shift acknowledges the vital part TQA has in upholding high standards in translation activities. These detailed guidelines, customized for various global areas, set key benchmarks crucial for preserving quality and uniformity in translations worldwide. In addition to these national and international standards, the translation industry has formulated various models. This paper examines models such as LISA QA, SAE J, MQM, and DQF, exploring the similarities and differences in quantifying translation quality within the industry.

The LISA QA model (The World Wide Web Consortium, n.d.), introduced by the Localization Industry Standards Association (LISA), employs an error deduction scoring system, assessing translation quality from both linguistic and

formatting perspectives, with specific parameters set for each. Linguistic assessment parameters include accuracy, mistranslation, terminology, style, and consistency, while formatting parameters cover layout, font, images, figure numbering, captions, and indexing. The model stratifies error severity into minor, major, and critical categories, each assigned specific weights, with total error scores calculated to assess translation quality. The model is adaptable, allowing for modifications in error categorization and severity, leading to varying outcomes. The implementation of this system in industrial practices demonstrates the shifting trend towards more organized and numerical methodologies within TQA.

The SAE J 2450 (Society of Automotive Engineers, n.d.), issued by the Society of Automotive Engineers (SAE), is primarily employed for quality assessment in the automotive industry and is applicable to manual, computer-assisted, and machine translation modalities. It, too, embraces an error scoring system, delineating errors into seven discrete categories: terminological, semantic, omissions, structural, spelling, punctuation, and others, each bearing differential weightings. Each error type is further divided into serious and minor categories. The assessment process involves five steps: 1) identifying errors within the target text, 2) categorizing the error into one of the seven types, 3) classifying the error as either “serious” or “minor”, 4) determining the numerical value of the error, 5) calculating a standardized score by adding all numerical weights and dividing by the number of words in the source language text. This standard quantifies translation errors, primarily focusing on linguistic form, with less emphasis on semantic quality. Widely used in the automotive industry and language service companies, it is particularly relevant in fields like medicine, industry, and manufacturing where terminological precision is crucial, effectively enhancing translation quality and efficiency.

The Dynamic Quality Framework (DQF) (The Translation Automation User Society, n.d.) developed by the Translation Automation User Society (TAUS) in 2011, offers a standardized method for assessing translation quality and supports quality management in translation and localization industries. The incorporation of DQF into industrial applications shows the increasing inclination toward flexible and tailored quality evaluation techniques, addressing the changing needs of contemporary translation endeavors. Combining various quality assessment models and tools, DQF is adaptable to different types of translation needs and projects. This framework adjusts dynamically based on the content type, purpose, and audience of the translation, exhibiting high flexibility and universality, catering to both human and machine translation. Based on LISA QA’s error classification, the DQF’s scoring criteria evaluate accuracy, linguistic form, terminology, genre, and prevalent error classifications in international standards (Görög, 2014), examining key dimensions such as accuracy, fluency, and regulatory compliance. This adaptability emphasizes the need for flexible evaluation systems capable of accommodating both human and machine translation processes.

The MQM (Multidimensional Quality Metrics) model ([Multidimensional Quality Metrics, n.d.](#)), developed by German Research Center for Artificial Intelligence (DFKI) and funded by the QTLaunchPad project of the EU, subsequently updated by QT 21. Parallels DQF, it offers an open and flexible system for assessing translation quality. The integration of MQM, especially its multi-faceted approach, mirrors a widespread industry transition towards more intricate and all-encompassing methods for evaluating quality. The MQM framework comprises four core elements: dimensions, issues, assessment methodologies, and weights, permitting customization to suit specific project requirements. MQM categorizes translation errors through eight core dimensions including accuracy, fluency, terminology, cultural adaptation, style, content authenticity, formatting, and localization, utilizing an error deduction system for scoring. Each error type is assigned varying weights, with the user granted the latitude to adjust these as required. The model categorizes error severity into four tiers: no errors, minor, major, and critical. To facilitate user-friendliness, MQM features a streamlined core model, simplifying the assessment procedure. Contrasting with the LISA QA model and SAE standards, MQM's model offers enhanced flexibility and objectivity, seamlessly integrating with prevailing models in the industry. Its wide applicability encompasses various fields and is equally suited for both human and machine-translated texts.

In an in-depth analysis of prevailing TQA models within the industry, it becomes apparent that these models predominantly employ an error deduction system as their fundamental evaluation approach. This approach intricately classifies translation errors, attributing distinct weights to varying error types, and subsequently calculates a comprehensive score for translation quality through weighted calculations. Such systems cater to the increasing demand for accountability and precision in translation quality evaluations, especially in critical sectors like legal, medical, and technical translations. This type of assessment methodology is equally suitable for assessing manually translated texts and evaluating the quality of machine translation outputs, as evidenced by models such as the DQF and MQM. The advent of models like MQM and DQF marks a shift in TQA frameworks towards more integrated, customizable, and flexible approaches. This evolution allows diverse users to tailor and implement quality assessment systems aligning with their specific requirements, target audiences, and application scenarios. For instance, a firm specializing in legal document translation might prioritize terminological accuracy and consistency, while an agency dealing with creative content may focus more on the adaptability of writing style and linguistic register. However, the heightened customization and flexibility inherent in these frameworks also introduce complexities in implementation, escalating the training demands on examiners who must not only possess translation proficiency but also grasp how to adeptly employ these assessment tools and models for continuous enhancement of translation quality. Additionally, with the rapid advancement of artificial intelligence and machine learning technologies, future TQA models may incorporate more intelligent features like au-

automatic error detection and real-time quality feedback, further augmenting the efficiency and outcomes of the translation process.

## 5. Conclusion

This study conducts a comprehensive exploration of the diverse methodologies, models, and operational practices in TQA across educational tests, certification exams, and the industry. At the heart of these sectors lies a unified set of assessment criteria: fidelity to the source text, effective delivery of the original information and intent, and adherence to the linguistic norms of the target language. These standards, tailored to meet the specific demands of various evaluative contexts, fields of application, and intended outcomes, manifest through a spectrum of practical approaches.

In translation education, the lack of universally recognized quantitative evaluation methods underscores an essential requirement for more systematic and objective assessment frameworks (Honig, 1997; Medadian & Mhabadi, 2015). Studies by Turner et al. (2010) and Koby and Baer (2005) have highlighted the efficacy of methodologies such as the ATA's error deduction system. With its detailed error categorization and transparent scoring approach, this system provides an invaluable tool for student self-enhancement and offers educators a definitive basis for unbiased assessment. This research supports the incorporation of industry-standard evaluation models, like the ATA's, into educational syllabi to improve teaching effectiveness and bolster students' understanding of translation nuances.

In certification and industry arenas, the widespread implementation of error deduction models reflects a commitment to methodical and quantitative translation analysis. This study recommends the ongoing refinement and personalization of these models to align with the dynamic nature of translation tasks, particularly those generated by artificial intelligence. Updating translation assessment criteria and error weightings is essential to align with technological progress, ensuring translations remain of superior quality in both linguistic accuracy and contextual relevance.

The swift advances in artificial intelligence and big data demand a reassessment of current TQA standards across education, certification, and industry. As the translation field increasingly integrates AI-generated texts, there is a pressing need to transition from traditional manual evaluations to more sophisticated, machine-assisted, or automated systems. This shift necessitates a comprehensive overhaul of assessment parameters, error weightings, and scoring methodologies, making sure they remain pertinent and efficacious amidst the evolving technological landscape and the ever-changing demands of the translation industry.

## Acknowledgements

This paper is part of the research project titled "Research on the Construction and Application of an Artificial Intelligence-Based Automatic Assessment System for Student Translation Quality" (No. 21YYC009), funded by the Beijing

Social Science Fund.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- American Translators Association (n.d.). *Framework for Standardized Error Marking*. <https://www.atanet.org/certification/how-the-exam-is-graded/error-marking/>
- Brunette, L. (2000). Towards a Terminology for Translation Quality Assessment: A Comparison of TQA Practices. *The Translator*, 6, 169-182. <https://doi.org/10.1080/13556509.2000.10799064>
- BS EN-15038 European Quality Standard (n.d.). *European Quality Standard EN-15038: 2006*. <http://qualitystandard.bs.en-15038.com/>
- Canadian Translators, Terminologists and Interpreters Council (n.d.). *Marking Scale*. <https://atio.on.ca/wp-content/uploads/2020/01/Marking-Scale-excerpt-EN.pdf>
- Chartered Institute of Linguistics (n.d.). *For Language and Linguists*. <https://www.ciol.org.uk/>
- Chen, Y. (2016). An Examination of Rater Performance on the Chinese-to-English Translation Section of TEM8 in 2015. *Technology Enhanced Foreign Language Education*, 171, 77-82.
- China Accreditation Test for Translators and Interpreters (n.d.). *Syllabus*. <https://www.catticenter.com/cattiyksgd>
- Colina, S. (2008). Translation Quality Evaluation: Empirical Evidence for a Functionalist Approach. *The Translator*, 14, 97-134. <https://doi.org/10.1080/13556509.2008.10799251>
- Görög, A. (2014). Dynamic Quality Framework: Quantifying and Benchmarking Quality. *Tradumàtica Tecnologies de la Traducció*, No. 12, 443-454. <https://doi.org/10.5565/rev/tradumatica.66>
- Government of Canada (n.d.). *Canada Language Quality Measurement System*. <https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-eng.html?lang=eng&i=1&index=alt&srchtxt=CANADIAN%20LANGUAG>
- Honig, H. G. (1997). Positions, Power and Practice: Functionalist Approaches and Translation Quality Assessment. *Current Issues in Language and Society*, 4, 6-34. <https://doi.org/10.1080/13520529709615477>
- House, J. (1981). *A Model for Translation Quality Assessment* (2nd ed.). Narr.
- House, J. (2002). Translation Quality Assessment: Linguistic Description versus Social Evaluation. *Meta*, 46, 243-257. <https://doi.org/10.7202/003141ar>
- Institute of Translation and Interpreting (n.d.). *ITI Code of Professional Conduct*. <https://www.iti.org.uk/resource/iti-code-of-professional-conduct.html>
- Koby, G. S., & Baer, B. J. (2005). From Professional Certification to the Translator Training Classroom: Adapting the ATA Error Marking Scale. *Translation Watch Quarterly*, 1, 33-45.
- Maneesriwongul, W., & Dixon, J. K. (2004). Instrument Translation Process: A Methods Review. *Journal of Advanced Nursing*, 48, 175-186. <https://doi.org/10.1111/j.1365-2648.2004.03185.x>
- Medadian, G., & Mahabadi, D. N. (2015). A Summative Translation Quality Assessment

- Model for Undergraduate Student Translations: Objectivity versus Manageability. *Studies about Languages*, No. 26, 40-54. <https://doi.org/10.5755/j01.sal.0.26.12421>
- Multidimensional Quality Metrics (n.d.). *What Is MQM*. <https://themqm.org/>
- National Accreditation Authority for Translators and Interpreters (n.d.). *NAATI Certification System*. <https://www.naati.com.au/certification/>
- National Education Examination Authority (n.d.). *College English Test—Band Four and Six (CET-4, CET-6) (Revised in 2016)*. <https://cet.neea.edu.cn/res/Home/1704/55b02330ac17274664f06d9d3db8249d.pdf>
- National Public Service Platform for Standards Information (n.d.). *Target Text Quality Requirements for Translation Services*. <https://std.samr.gov.cn/gb/search/gbDetailed?id=F159DFC2A8DD47EFE05397BE0A0AF334>
- Pym, A. (2010). *Translation and Text Transfer—An Essay on the Principles of Cross-Cultural Communication* (2nd ed.). Intercultural Studies Group.
- Si, X. (2004). On the Research of Translation Quality Assessment Model from the Perspective of Functional Linguistics. *Foreign Language Education*, 25, 45-50. <https://doi.org/10/gnnxxf>
- Si, X. (2008). A Re-Examination of the Quality Assessment Model for Translated Parts of Speech—A Functional Linguistic Approach. *Chinese Translators Journal*, 2, 57-60.
- Society of Automotive Engineers (n.d.). *SAE Standards*. <https://www.sae.org/standards>
- The Translation Automation User Society (n.d.). *Dynamic Quality Framework*. <https://www.taus.net/resources/blog/category/dynamic-quality-framework>
- The World Wide Web Consortium (n.d.). *The Operating Objectives and Activities of the Localisation Industry Standards Association*. <https://ot2009.files.wordpress.com/2009/05/5-lisa-best-practice-guide.pdf>
- Turner, B., Lai, M., & Huang, N. (2010). Error Deduction and Descriptors—A Comparison of Two Methods of Translation Test Assessment. *Translation & Interpreting*, 2, 11-23.
- Wang, J., & Zhu, Z. (2016). Development and Application of Automated Chinese-English Translation Scoring System. *Journal of Yangzhou University (Humanities & Social Science)*, 20, 123-128. <https://doi.org/10.19411/j.cnki.1007-7030.2016.02.018>
- Wu, G. (2007). Progress, Meta-Assessment and Direction of Development of Contemporary Models of Quality Assessment for Chinese and Western Translators. *Foreign Language Research*, 4, 123-128.
- Yang, Z. (2012). Quantitative Assessment of Translation Quality: Models, Trends, and Implications. *Foreign Language Research*, No. 6, 65-69. <https://doi.org/10.13978/j.cnki.wyyj.2012.06.013>
- Zhao, D., & Cai, Y. (2018). The Impact of the Translators' Humanistic Knowledge on the Quality of Translation—In the Case of TEM8 Translation in 2018. *Foreign Language Testing and Teaching*, 2, 20-24, 31.