

## A critical perspective on the translation quality assessments of five translators organizations: ATA, CTTIC, ITI, NAATI, and SATI

Mehmet YILDIZ<sup>1</sup>

**APA:** Yıldız, M. (2020). A critical perspective on translation quality assessments of five translators' organizations: ATA, CTTIC, ITI, NAATI, and SATI. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (18), 568-589. DOI: 10.29000/rumelide.706390

### Abstract

The present paper discusses translation quality assessments by adopting a critical perspective on five translators organizations, which are intended to assess the quality of non-literary translations with a particular focus on their objectivity, validity, and inter-rater reliability. Within this framework, it aims to contribute to the related literature (1) by discussing the objectivity, validity, and inter-rater reliability of the quality assessment methods of five translators organizations, namely the American Translators Association (ATA), the South African Translators' Institute (SATI), the Canadian Translators, Terminologists and Interpreters Council (CTTIC), the Institute of Translation and Interpreting (ITI in UK), and the National Accreditation Authority for Translators and Interpreters (NAATI in Australia), and (2) by making suggestions on quality assessment concerning non-literary translation in view of the findings hereof. The study qualitatively analyzes the content of the guides and rubrics provided on the websites of these five organizations and discusses the objectivity, validity, and inter-rater reliability of their assessments in consideration of seven parameters, i.e. "purpose of assessment", "purpose of assigned translation", "duration", "source text", "assessor", "marking", and "grading". The findings showed that each organization suffers from varying degrees of objectivity, validity, and inter-rater reliability issues.

**Keywords:** Translation quality assessment, objectivity, validity, inter-rater reliability, translators organizations.

### Beş çevirmenler derneğinin çeviri kalitesi ölçme uygulamalarına eleştirel bir bakış: ATA, CTTIC, ITI, NAATI ve SATI

#### Öz

Bu çalışma, beş çeviri örgütünün yazınsal olmayan çevirilerin kalitelerini değerlendirmeyi amaçlayan çeviri kalitesi değerlendirme uygulamalarını; nesnellik, geçerlilik ve değerbiçiciler arası güvenilirlik açısından eleştirel bir bakış açısıyla değerlendirmektedir. Bu bağlamda, (1) the American Translators Association (ATA), the South African Translators' Institute (SATI), the Canadian Translators, Terminologists and Interpreters Council (CTTIC), the Institute of Translation and Interpreting (ITI - Birleşik Krallık), ve the National Accreditation Authority for Translators and Interpreters (NAATI - Avustralya)'dan oluşan beş örgütün çeviri kalitesi değerlendirme yöntemlerini nesnellik, geçerlilik, ve değerbiçiciler arası güvenilirliklerini tartışarak ve (2) bu çalışmanın bulguları çerçevesinde yazınsal olmayan çevirilerin kalitelerinin değerlendirilmesine değgin önerilerde bulunarak alanyazına katkı sağlamayı amaçlamaktadır. Çalışma; söz konusu örgütlerin internet sitelerinde

<sup>1</sup> Dr. Öğr. Üyesi, Çanakkale Onsekiz Mart Üniversitesi, Fen Edebiyat Fakültesi, Batı Dilleri ve Edebiyatları Bölümü (Çanakkale, Türkiye), mehmetyildiz@comu.edu.tr, ORCID ID: 0000-0001-9482-4358 [Makale kayıt tarihi: 14.10.2019-kabul tarihi: 20.03.2020; DOI: 10.29000/rumelide.706390]

verilen kılavuz ve standartların içeriklerini nitel olarak çözümlenmekte ve “deęerlendirmenin amacı”, “istenilen çevirinin amacı”, “süre”, “kaynak metin”, “deęerlendirici”, “puanlama” ve “derecelendirme”den oluřan yedi deęiřkeni göz önünde bulundurarak bu beř örgütün nesnellik, geçerlilik ve deęerbiçiciler arası güvenilirliklerini ele almaktadır. Bulgular, herbir örgütün deęiřen düzeylerde nesnellik, geçerlilik ve deęerbiçiciler arası güvenilirlik sorunları olduęunu göstermiştir.

**Anahtar kelimeler:** Çeviri kalitesinin ölçülmesi, nesnellik, geçerlilik, deęerbiçiciler arası güvenilirlik, çeviri örgütleri.

## 1. Introduction

The primary concern of the present paper is to critically discuss the objectivity, validity and inter-rater reliability of translation quality assessments of the American Translators Association (ATA), the South African Translators' Institute (SATI), the Canadian Translators, Terminologists and Interpreters Council (CTTIC), the Institute of Translation and Interpreting (ITI in UK), and the National Accreditation Authority for Translators and Interpreters (NAATI in Australia) – and to make some suggestions to contribute to the available literature on the assessment of non-literary translations.

The question of objectivity in translation evaluation is not a recent topic. In his manifesto-like paper, *The Name and Nature of Translation Studies*, James Holmes (2000) lists translation criticism, along with translation teaching, translation aids and translation policy, as one of the areas of applied translation studies. This indicates that translation-oriented judgments have always been deemed to constitute an integral part of translation studies and therefore should be investigated as frequently as possible to improve practices both in industry and academia. According to Holmes (2000, 182) it is doubtless that,

The activities of translation interpretation and evaluation will always elude the grasp of objective analysis to some extent, and so continue to reflect the intuitive, impressionist attitudes and stances of the critic. But closer contact between translation scholars and translation critics could do a great deal to reduce the intuitive element to a more acceptable level (Holmes 2000, 182).

This shows that the dichotomy of ‘subjectivity vs. objectivity’ in translation assessment has been a pivotal concern of translation studies since the very first years of its emergence as a discipline. The excerpt also highlights that translation scholars and translation critics should be in touch to minimize subjectivity in translation assessment and promote objective practices. Therefore, this paper can be considered as a scholarly attempt to provide translation critics, assessors herein, with feedback on their evaluations.

Nicole Martínez Melis and Amparo Hurtado Albir (2001, 283) go beyond the aforementioned dichotomy and suggest a tripartite body of translation assessment; intuitive, partial, and reasoned. “Intuitive assessment is subjective, impression-based and does not follow explicit criteria”. “Partial assessment does not take into account the sum total of factors involved in a translation”. “[Reasoned assessment] is objective, using scales established on the basis of objective criteria which define and assign a value to the error type. Its apparent reliability would seem to indicate that it is the most appropriate type of assessment in all areas of translation” (Martinez-Melis and Albir 2001, 283).

The reader can realize that the first two would fail to provide the same or similar results each time they are administered even when the same condition is provided and to cover the very distinctive characteristics of a translational phenomenon by being either superficial or partial, thus not measuring what it is intended to measure. To Martinez-Melis and Albir's categorization, a fourth assessment type

can be added: intersubjective assessment. Intersubjective is defined as “verifiable by several subjects/persons” by Momka Krein-Kutiile (2003: 38). Accordingly, if an assessor discusses his/her judgment with other assessors, then “the process can become intersubjective” (Hansen 2003, 36).

The following four sections discuss translation assessment as a subjective, objective, intersubjective, and idiocentric and allocentric phenomenon to provide a contained theoretical framework to the descriptive analysis of the corpus consisting of the data available on the websites of ATA, SATI, CTTIC, ITI, and NAATI. In “Method”, the research questions, the corpus, and the method of analysis are provided. In “Findings”, the data obtained from the analyzed corpus are presented in consideration of seven parameters, i.e. (1) purpose of assessment, (2) purpose of assigned translation, (3) duration, (4) source text, (5) assessor, (6) marking, and (7) grading. In the last section, the drawn conclusions are offered and some relevant suggestions are made as to how to produce a more valid, reliable, and objective translation quality assessment tool or to improve an existing one.

## **2. Translation quality (assessment) as a subjective phenomenon**

Translation quality assessment has been considered a subjective phenomenon for two reasons in particular: method of assessment (Morin et al. 2017, 207) and assessors' varying rubrics of ‘acceptable’ text (Bowker 2000, 183). The latter suggests that quality too is relative since “the notion of quality has very fuzzy and shifting boundaries” (Bowker 2000, 183) and “standardization of quality is thus a fuzzy grey area” (Al-Qinai 2000, 498). Subjective quality and “how to express quality” (Rothe-Neves 2002, 114) are listed among the main problems of quality assessment, especially “if it is taken to involve individual and externally motivated value judgment alone” (House 2001, 255) because “quality is not ‘objectively’ given, but depends on the text user and his/her assessment criteria” (Schäffner 1997, 2). This last remark brings the reader back to the role of evaluator. In this looping conceptualization, a translation-focused description of quality will prove greatly contributory to the discussion of translation quality and its assessment.

The dictionary definition of quality is “the standard of something as measured against other things of a similar kind; the degree of excellence of something<sup>2</sup>”. It can be derived from this definition that quality is a ‘thing’ whose presence or absence can only be appreciated when and if it is ‘measured against other things of similar kind’ and its existence absolutely relies on external criteria, which constitute the primary concern of this study. Thus, “the meaning of quality can vary considerably for different individuals, groups, and contexts” (Castilho et al. 2018, 11).

Geoffrey Koby et al. (2014, 416-417) offer “two contrasting definitions” of quality translation: a broad definition and a narrow definition. The former reads: “A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs”. On the other hand, the narrow definition is “one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture”. In the first definition, the reader may sense the emphasis on the functionality of a target text and the influence of the functional paradigm because the definition is roughly based on the formulation of the target text around the purpose and specifications presented by the commissioner.

2 <https://en.oxforddictionaries.com/definition/quality>, accessed 07 October 2018

This means that his/her expectations will determine the quality of the target text, where minor errors may be acceptable as long as the target text fulfils what it has been produced for. However, in the latter definition ‘loyalty to source text’ and ‘observation of cultural norms’ are highlighted, which dictates that no error is tolerable and the acceptability of the target text is tested against culture, which is a collective production rather than an individual request. Thus, “the choice of one definition of translation quality over the other will influence the framework for developing translation quality metrics and have practical consequences for measuring translation quality” (Koby et al. 2014, 417).

Gyde Hansen’s compilation of quality definitions (2008, 206) epitomizes how diversely quality can be defined and how influential some specific translation paradigms are in the way quality is perceived:

1. Quality is a question of *individual perception*<sup>3</sup>.
2. Quality is a *cultural issue*.
3. Quality is *meeting the clients’ needs*.
4. Quality is *fulfillment of the skopos*.
5. Quality is *fitness for use*.
6. Quality is *the degree of equivalence between source and target text*.
7. Quality is *the result of a good process*.
8. Quality is also described as *not merely an absence of errors* (Hansen 2008, 260).

To elaborate, quality may be theoretically defined in countless ways because each individual would tend to resort to assorted criteria to evaluate a translated text. Besides, quality is culture-dependent; in other words, the way it is conceptualized may vary across different cultures, societies, and even economies. From a functionalist viewpoint, the primary benchmarks of judgment about translation quality is to what extent commissioner’s expectations are satisfied, the purpose provided by him/her is achieved, and the produced translation addresses his/her need(s). Furthermore, quality might be simply characterized by the degree of ‘faithfulness’ to the source text from an equivalence-oriented linguistic perspective while a possible definition may transcend faithfulness and may adopt a different benchmark to refer to quality instead of reducing a ‘high-quality’ translation to an error-free textual rendition. Moreover, a process-oriented definition might well be built on the procedural aspects of translation act by surmising that ‘a good process would yield a good translation’.

These definitions show that any effort to evaluate translation quality adopts a theoretical or ideological stance, be it intentionally or unintentionally, because “evaluating the quality of a translation presupposes a theory of translation. Thus different views of translation lead to different concepts of translational quality, and hence different ways of assessing it” (House 2009, 222).

Even though “passing any ‘final judgment’ on the quality of a translation that fulfills the demands of scientific objectivity is very difficult (House 2001, 255), and subjectivity is inevitable (Zehnalová 2013, 43; Bowker 2000, 183), the inevitable subjectivity “does not invalidate the objective part of the assessment” (House 2001, 256). This is because an objective assessment characterized by a valid and reliable method of evaluation is sought to avoid subjective judgments. Acknowledging the unavoidability of subjectivity in translation quality assessment does not necessarily presuppose that objectivity is

3 Original italics from 1 through 8.

unattainable to a certain extent. Accordingly, the following section discusses why objectivity matters for translation quality assessment.

### **3. Translation quality assessment as an objective phenomenon**

Achieving an objective assessment of translation quality is difficult (House 2001, 255; Bittner 2011, 76) and may be even impossible (Williams, 1989: 14; Zehnalová 2013, 42). Yet because delivering a fair judgment concerning the quality of a target text entails down-to-earth standards, “the assessment of the quality of a translation should ideally be based on more objective criteria” (Bittner 2011, 76). These remarks hint that absolute objectivity is unlikely to occur; nevertheless, if an evaluator wishes to go beyond prescriptively labeling a translation “good” or “bad”, then he/she should apply criteria that are as objective as possible. This is why “both the language industry and translation studies urgently need a method to measure translation quality as objectively as possible” (Koby et al. 2014, 415-416) and, as Malcolm Williams (2009) states, although “national and international translation standards now exist” to meet the need for objective rubrics, “there are no generally accepted objective criteria for evaluating the quality of translations” (Williams 2009, 3). “Because the evaluation of translations can be carried out for many distinct reasons, different criteria and factors take on varying degrees of importance” (Bowker 2000, 183), which makes it impossible to develop an overarching body of criteria and factors

It would be too naïve to ask for a set of evaluative criteria applicable to any translation situation. Nonetheless, there are a number of prominent characteristics of an objective assessment. Validity and reliability are two of these integral features. The former is the quality of an assessment tool that ensures that it measures what it intends to measure, while the latter can be defined as an assessment tool’s ability to produce consistent results, i.e. the same results under the same conditions. Researchers should remember that an assessment tool cannot be valid by compromising reliability but can be reliable without being valid. In other words, a tool’s yielding the same results even when employed in countless administrations with the same characteristics refers to its high reliability, which does not necessarily guarantee that it measures what it is supposed to measure.

The significance of validity and reliability being said, the problems likely to appear during the production and application of a translation quality assessment tool should be discussed. For Williams (2009, 5), these problems are associated with the evaluator, level of target language rigor, seriousness of errors of transfer, sampling, quantification of quality, levels of seriousness of error, multiple levels of assessment, and the purpose or function of translation quality assessment. Evaluators are active participants in an assessment process and are expected to have the required linguistic and extralinguistic knowledge. They are also expected to set an acceptability filter to decide what to count and not count as an error, whether to categorize an error as a linguistic or transfer error, and how serious an error is. Most assessment methods are error-based because simultaneous assessment of translation solutions and translation errors takes a considerable amount of time. This is the reason why sometimes sampling a small portion of a text may prove fruitful, yet assessors should make sure that the excerpt is representative enough to allow the final score to be generalizable over the entire target text. Error-based assessment is a micro-analytical approach and “has been used extensively not only because it saves time but also because it provides error counts as a justification for a negative assessment” (Williams 2009, 6). Juliane House (2001, 255) proposes that an evaluator should adopt a multi-perspective viewpoint by “flexibly mov[ing] from a macro-analytical focus to a micro-analytical one, from considerations of ideology, function, genre, register, to the communicative value of individual linguistic items” and that an attitude of this kind will enable him/her to be as objective as possible (House 2001, 255). However, Williams (2009, 7)



expresses his concerns over the operationalizability of these macro-level parameters along with micro-level ones and the generatability of “an overall quality rating for the translation” in view of these two levels and also states that the purpose and function of assessment should be well defined because where and why an assessment tool is employed will determine the tool’s characteristics. This is why some pre-production and pre-administrational decisions should be made to fend off validity-related concerns. Will it be administered at a professional organization? Has the target text as the assessment subject been produced by a professional? For whom has the target text been produced? Is the translated text a literary rendition? Or a non-literary text? Such macro-considerations will be significantly influential in the production and administration of an assessment tool and certainly in the evaluation of the data obtained with the tool.

It seems that objectivity is considered an indispensable evaluation component when assessing the quality of non-literary translations and requires devoted adherence to a set of preset criteria produced in consideration of the purpose a translated text presumably fulfills and the fact that the translated text should meet some minimum universal standards of accuracy.

Marcel Thelen (2008, 419) puts forward three generic considerations pertaining to translation quality assessment: (1) formal and semantic equivalence between source and target text, (2) consideration of “the locale of the TT’s readers”, and (3) accuracy and naturalness of the target language used. This account requires compliance with linguistic and extralinguistic aspects, including but not limited to structural and meaning-based correspondence, target culture and reader, and naturalness of the target language use.

Martinez-Melis and Albir (2001, 283) list four applicable principles “to carry out a fair and objective assessment”: (1) The evaluator should comply with some parameters and let the evaluatee know them. (2) The parameters should be determined based on the context (e.g., professional translation) and function of the assessment. (3) What is to be assessed and at what level the assessment is to be performed should be considered. The evaluator should also be aware of his/her capacity to assess. (4) The evaluator is also expected to be knowledgeable about the indicators of the intended phenomenon.

These remarks from the evaluator’s viewpoint are closely related to the construct validity<sup>4</sup> of a proposed assessment tool, which “refers to the extent to which a test measures what it is designed to measure” (Angelelli 2012, 175). Supposedly objective assessors should know about the parameters to follow as they run the assessment and about the indicators of a translational phenomenon. The extent to which an assessor associates these parameters with the observed phenomena will determine the degree of construct validity of an assessment (tool). Additionally, an assessor’s knowledge about translational phenomena and assessment criteria and awareness of the context and purpose of a target text should improve the intra-rater reliability of the assessment. “Scoring is ideally done based on fixed and objective criteria. Each instance of scoring by a grader should be similar to other instances of scoring that the same grader performs. This quality is known as intra-rater reliability” (Angelelli 2009, 18). Therefore, providing assessors with criteria to follow and training them on these parameters should increase the likelihood that the assessors will repeatedly apply the same criteria in future grading activities and their assessments will yield the same/similar results under the same/similar conditions.

4 Considered as a “catch-all” concept overarching other two prominent validity types, i.e. content and criterion validity (Messick, 1994: 3, Messick, Samuel, Validity of Psychological Assessment: Validation of Inferences from Persons’ Responses and Performances as Scientific Inquiry into Score Meaning).

Some concepts stand out in the lists above, which are evaluator, evaluatee, criteria, context, function, assessment object, evaluator's competence, and indicators of the observed phenomenon. As the active agents of the evaluation process, evaluators can be teachers, clients, and professional assessors and the quality of a translation is tested against the criteria they have produced or adopted. Thus, an evaluator's attitude towards the notion of evaluation and its constituent elements can be expected to determine its quality; however, in a supposedly objective evaluative effort, this central position of an evaluator should be counterbalanced by the context and function of evaluation. Martinez-Melis and Albir (2001) state that the assessment context should incorporate published translations, professional translation, and translation teaching, and that assessment may fulfill a summative, diagnostic, and/or formative function (Martinez-Melis and Albir 2001, 283). Even though technical texts and their translations are also exploited for training purposes, they typically occur in professional settings; therefore, their assessment is expected to assume a summative rather than a diagnostic and formative function because summative assessment is conducted on a finished product, here a non-literary translation, and "provides evidence for decision-making" as to whether the assessee is competent enough in the rendition of a translation in a particular specialized field "to proceed to next unit", "to be awarded certification", or furnished "with a professional qualification" (Hatim and Mason 2005, 166). Since the present study primarily focuses on translation quality assessment in non-literary translation, summative assessments in professional translation settings have been included in the corpus.

The foregoing remarks indicate that certain criteria should be predetermined and provided with justifiable and reasoned content in order for assessors to be able to operationalize them objectively and effectively. Yet the reader can realize that this proposition raises another critical issue: What if an assessor (or each of two or more assessors) applies the criteria differently? This is why the results should be intersubjectively verified. In this context, the following section discusses what intersubjectivity is and why it matters.

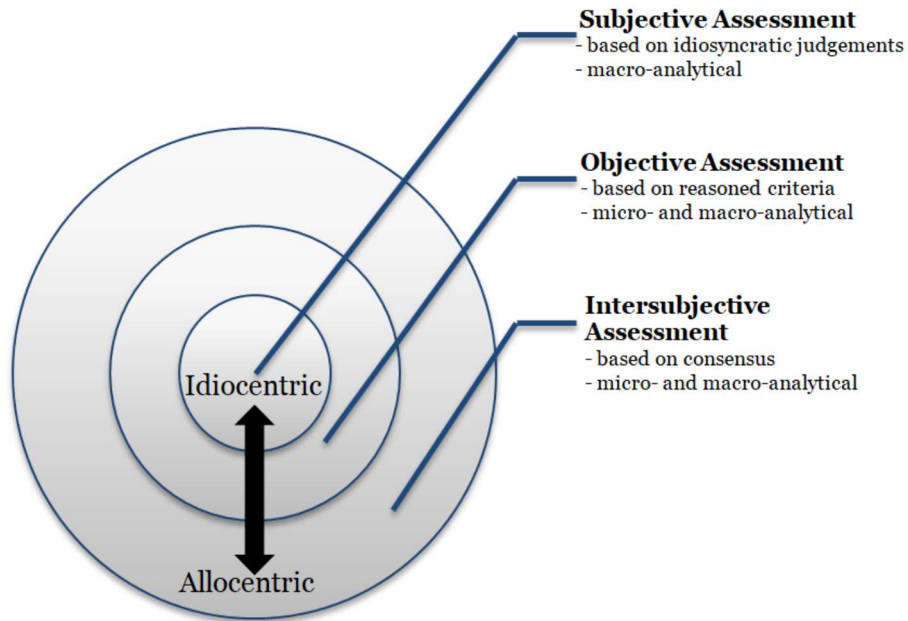
#### **4. Translation quality assessment as an intersubjective phenomenon**

Objectivity might turn out to be a relative concept if based upon judgments of a single assessor applying the criteria that are supposed to help administer an objective assessment. Hence, a supposedly objective assessment should be buttressed by intersubjective evaluative practices. Thus assessors "have to make explicit the grounds for [their] judgment basing it on a theoretically sound and argued set of intersubjectively verifiable [...] procedures (House 2001, 256). The grounds for judgment based on a theoretically sound set of procedures can be anticipated to boost the validity of a tool or an assessment process since such theoretical substantiation is supposed to help an assessment tool take into account already available scholarly knowledge and to allow it to target the intended translational phenomenon relevant to the evaluative process. Moreover, intersubjective verifiability should increase the level of validity and inter-rater reliability because overlapping judgments made by two or more assessors will show the effectiveness of the criteria selection and will indicate that assessors who might make subjective decisions based on their experience and in consideration of their ideal target text are properly applying the selected criteria. This could indicate that the assessment process/tool is likely to yield the same results if and when employed under the same conditions, which refers to its high interrater reliability. Iris Schrijver, Leona Van Vaerenbergh, and Leona Van Vaerenbergh (2012, 116) claim that "subjectivity can be controlled if interrater reliability is high" and Arle Lommel, Maja Popović, and Aljoscha Burchardt (2014, 33) state that "demonstrating a high degree of [inter-rater reliability] is a necessary step to showing that an assessment metric is reliable". It can be thus asserted that intersubjectivity substantially matters for an assessment tool's interrater reliability and validity because

it makes it possible to show whether the tool will produce the same/similar results related to the intended phenomenon when it is synchronically and diachronically administered by two or more assessors. Considering translation assessment as a subjective, objective and intersubjective phenomenon, the following section presents a new classification of translation quality assessment: idiocentric and allocentric assessment.

### 5. Translation quality assessment as an idiocentric and allocentric phenomenon

Three characteristics of human-centered assessments, i.e. subjective, objective, and intersubjective, are discussed above. It is clear that subjective assessment is characterized by idiosyncratic and intuitive judgments. Subjective assessors generally employ “anecdotal approaches to the evaluation of translations” (House 1997, 3). Such an evaluative technique is “impression-based” (Martinez-Melis and Albir 2001, 283) and therefore builds on macro-analysis which lends itself to qualitative assessment rather than evaluative efforts relying on objective quantification. Although impression-based evaluations are highly revered in literary discussions, where interpretative appreciation of a literary translated work is central, it should be avoided in the quality assessment of non-literary translations particularly since non-literary texts are not produced for aesthetic appreciation. This is why its quality should be objectively and preferably intersubjectively tested against efficient parameters to ensure that it is capable of “transmitting the message faithfully” (Yazıcı, 2007: 56).



**Fig. 1.** Three characteristics of human-centered translation quality assessment

Fig 1. summarizes subjective, objective and intersubjective assessment in terms of the source of judgment (assessor himself/herself, criteria provided, or two or more assessors) and level of assessment (micro- or macro-level) and introduces two new assessment categories (idiocentric and allocentric assessment) as to whether the assessment is based on a personal or a collective judgment. Assessors of non-literary translations are expected to refrain from idiocentric evaluations but to act allocentrically. “Allocentrism is defined as greater emphasis on the views, needs, goals and concerns of the ingroup than of oneself. Idiocentrism is greater emphasis on own views, needs, goals and concerns than on the views, needs, goals and concerns of others” (Triandis 1983, 16). To put it into context, allocentric assessors



prioritize externally produced standards, as in assessment rubrics, and judgments of other assessors, as in an assessment committee, over their intuitive, impression-based, and anecdotal evaluations. Therefore, it can be claimed that as assessment approaches the outer rim of the circle in Fig. 1, it becomes more reliant on externally provided parameters or more interdependent on other assessing actors, and thus more allocentric.

## 6. Method

### 6.1. Research questions

The foregoing discussion on three characteristics of translation quality assessment provides a number of parameters that will serve as evaluative guidelines and allow the author of the present paper to apply in order to analyze the assessment tools for non-literary translation available in the translation/language market. Following are the research questions produced based on these parameters:

- A. Why is the assessment administered?
- B. Who administers the assessment?
- C. Does the assessment have some subjectivity or objectivity-related issues?
- D. May the assessment cause some validity or inter-reliability concerns?
- E. Is the assessment idiocentric or allocentric?
- F. Is the assessment error-based or solution-based?
- G. Are errors and/or solutions adequately operationalized for the purpose of the assessment? (e.g., linguistic error vs. translation error, grading, seriousness of errors)

### 6.2. Corpus

Because the study is intended to investigate professional evaluative practices intended to assess translation quality in non-literary translation, the corpus consists of the assessment guides and rubrics of ATA, SATI, CTTIC, NAATI, and ITI, used to regulate the assessment process (Table 1). Only the guides and rubrics available on these organizations' websites were included in the corpus.

These organizations were selected for the following reasons:

- a. They are "prominent" translators organizations with translation exams whose rubrics and guides are available online.
- b. They are located on four different continents, namely America, Africa, Europe, and Australia, which is believed to promote the representativeness of the sampled data.
- c. They are located in the countries with English as their native language, which helped the author confidently analyze his corpus – and the related websites if need be – because his working language pair is Turkish-English.

	<b>Guides/Rubrics</b>
<b>ATA</b>	ATA Certification Program Framework for Standardized Error Marking – Version 2017 <sup>5</sup>
	Flowchart for Error Point Decisions – Version 2009 <sup>6</sup>
	A Guide to the ATA Certification Program <sup>7</sup>
	Computerized Certification Exam Guide <sup>8</sup>
	Procedure to Appeal Certification Exam Review <sup>9</sup>
<b>SATI</b>	Information Regarding the Accreditation Examinations <sup>10</sup>
	Accreditation Examinations: Guidelines on Marking <sup>11</sup>
	Framework for Marking of Translation Accreditation Examinations <sup>12</sup>
<b>CTTIC</b>	Candidate’s Guide for the CTTIC Standard Certification Examination in Translation <sup>13</sup>
	CTTIC Standard Certification Translation Examination: Marker’s Guide <sup>14</sup>
<b>NAATI</b>	Policy for the Use of Equipment and Reference Materials in Certified Translator Tests and Certified Advanced Translator Tests <sup>15</sup>
	Certified Translator Test: Instructions for Candidates <sup>16</sup>
<b>ITI</b>	Applicant Handbook: Assessment Guide for Translators <sup>17</sup>

**Table 1.** Assessment guides and rubrics of the organizations

### 6.3. Data analysis

The present study builds on content analysis to analyze the guides and rubrics of ATA, SATI, CTTIC, NAATI, and ITI. “Content analysis focuses on analyzing and interpreting recorded material to learn about human behavior. The material may be public records, textbooks, letters, films, tapes, diaries, themes, reports, or other documents. Content analysis usually begins with a question that the researcher believes can best be answered by studying documents” (Ary, Jacobs, and Sorensen 2010, 29-30). The recorded materials used herein are guides and rubrics available on these five organizations’ websites and were studied to reveal the evaluative practices at them in consideration of the research questions above since among the purposes of content analysis is “to describe prevailing practices” (Ary, Jacobs, and Sorensen, 2010, 457). This analysis type “can be undertaken with any written material” (Cohen, Manion, and Morrison 2000, 245. “Content analysis takes texts and analyzes, reduces and interrogates them into summary form through the use of both pre-existing categories and emergent themes in order to generate or test a theory” (Cohen, Manion, and Morrison 2000, 476). The analysis in the present article was constructed based on seven parameters isolated from the analysis of the corpus, which are (1) purpose of assessment, (2) purpose of assigned translation, (3) duration, (4) source text, (5) assessor,

- 5 [http://www.atanet.org/certification/Framework\\_2017.pdf](http://www.atanet.org/certification/Framework_2017.pdf)  
6 [http://www.atanet.org/certification/aboutexams\\_flowchart.pdf](http://www.atanet.org/certification/aboutexams_flowchart.pdf)  
7 [http://www.atanet.org/certification/aboutcert\\_overview.php#1](http://www.atanet.org/certification/aboutcert_overview.php#1)  
8 [http://www.atanet.org/certification/aboutexams\\_computerized.php](http://www.atanet.org/certification/aboutexams_computerized.php)  
9 [http://www.atanet.org/certification/aboutreview\\_appeal.php](http://www.atanet.org/certification/aboutreview_appeal.php)  
10 [http://translators.org.za/sati/cms/downloads/dynamic/sati\\_accreditation\\_general\\_guidelines\\_english.pdf](http://translators.org.za/sati/cms/downloads/dynamic/sati_accreditation_general_guidelines_english.pdf)  
11 [http://translators.org.za/sati/cms/downloads/dynamic/majors\\_minors\\_english.pdf](http://translators.org.za/sati/cms/downloads/dynamic/majors_minors_english.pdf)  
12 [http://translators.org.za/sati/cms/downloads/dynamic/markings\\_guidelines\\_translation\\_english.pdf](http://translators.org.za/sati/cms/downloads/dynamic/markings_guidelines_translation_english.pdf)  
13 [http://www.cttic.org/examDocs/CandidateGuide\\_En1011.pdf](http://www.cttic.org/examDocs/CandidateGuide_En1011.pdf)  
14 <http://www.cttic.org/examDocs/guide.markersE.pdf>  
15 <https://www.naati.com.au/media/1952/policy-use-equipment-reference-materials-trans-tests.pdf>  
16 <https://www.naati.com.au/media/1951/ct-candidate-instructions-finalpdf.pdf>  
17 <https://www.iti.org.uk/attachments/article/992/ITI%20Applicant%20Handbook%20v4.pdf>

(6) marking, and (7) grading. These themes are discussed below under four headings because some related themes were clustered.

## 7. Findings

This section presents the findings in consideration of seven parameters, i.e. (1) purpose of assessment, (2) purpose of assigned translation, (3) duration, (4) source text, (5) assessor, (6) marking, and (7) grading. The related themes were merged; hence, they were discussed under four headings. The data in the tables below were isolated from the guides and rubrics in Table 1.

### 7.1. Purpose

*7.1.1. Purpose of assessment.* The question of why a quality assessment tool is employed substantially influences its production and administration because an assessment may assume a diagnostic, formative, or summative purpose as well as an educational or professional aim.

	ATA	SATI	CTTIC	ITI	NAATI
<b>Purpose</b>	To assess an individual's ability to provide quality, professional translation in a specified source-target language combination.	To test the final translation product that the candidate can present.	Intended for experienced translators who wish to have their competence recognized by their peers.	A compulsory entry test for professional translators wishing to gain qualified membership to ITI.	A performance-based test of your translation or interpreting skills

**Table 2.** Purposes of the organizations' assessments

Table 2 shows that while ATA, SATI, CTTIC, and NAATI administer tests to award their applicant translators a certification and vest them with a right to avail themselves of the accompanying advantages, ITI offers it as a membership test to grant assesseees a "qualified membership". As Basil Hatim and Ian Mason (2005, 166) put it, because a summative evaluation "provides evidence for decision-making" as to whether the assessee is competent enough in the rendition of a translation in a particular specialized field "to proceed to next unit", "to be awarded certification", or furnished "with a professional qualification", these exams can be labeled as summative tests.

*7.1.2. Purpose of translation task.* Besides having a purpose of its own, a translation quality assessment test should also provide the assessee with a purpose for him/her to be able to make macro-decisions concerning the translation task at stake, so that the assessee can supposedly determine a guideline to follow throughout the translation process. The purpose presented with the translator also helps the assessor pass grounded rather than intuitive judgments by serving as a benchmark against which translation decisions are tested. Without such a standard the assessor will have to come up with his/her own benchmarking system which may lead to subjective evaluations.

	ATA	SATI	CTTIC	ITI	NAATI
<b>Translation brief</b>	Translation Instructions, specifying the context within which the translation is to be performed (text source and translation purpose, audience, and medium)	Instructions for specific language combinations will accompany each paper. In the case of sworn translations a set of guidelines will be included with each paper.	N/A	N/A	Translation Brief for domain and type of text, information about the source text, including author or issuing authority, and target audience and purpose of the translation

**Table 3.** Purposes of translation tasks

Table 3 reveals that whereas ATA, SATI, and NAATI offer translation instructions along with each source text, assesseees are not furnished with a translation brief by CTTIC and ITI, to the best knowledge of the author. Such a practice may lead to some validity- and objectivity-related concerns. A translator not provided with a brief has to find a purpose to translate, which might result in the rendition of different target texts from what is acceptable by assessors because the assessee' macro-perspective may not be congruent with that of the assessor. In this case, the criteria applied by the assessor may partially cover the intended translational phenomena, which is a risk that the assessors of CTTIC and ITI may have to face.

### 7.2. Source text and duration

The translation associations whose assessments are analyzed request assesseees to translate texts of varying lengths in varying durations. This may present some data as to the level of their ecological and construct validity in that a relatively long duration allowed for a relatively short text may not yield the data concerning the central strategic competence which requires superior time management and governs the translation process to allow translators to quickly render translations by effectively and efficiently operationalizing the needed skills.

ATA, CTTIC, and NAATI let candidates finish the assigned translation task in three hours, three hours, and three and half hours, respectively (Table 4). In these durations of approximately the same length, two source texts with different numbers of words are presented to assesseees. ATA and NAATI ask assesseees to choose two out of three texts. CTTIC too requires its candidates to translate two source texts, i.e. one compulsory and one elective out of two texts. The number of words in the source texts ranges from about 225 to 275 words in ATA's exam and from about 175 to 185 words in CTTIC's and amounts to about 250 words in NAATI's, in which the assessee is also required to revise a non-specialized translation task (Table 4). In other words, translators in ATA's assessments are supposed to spend three hours for two source texts, each with an average of 250 words, while the CTTIC's exam obligates translators to translate two texts, each with an average of 180 words, in three hours. On the other hand, NAATI requires its assesseees to complete the translation of two texts, each with approximately 250 words, and the revision of a translation task with around 250 words, in three and half hours.

	ATA	SATI	CTTIC	ITI	NAATI
<b>Source Text</b>	Two out of three texts with about 225 to 275 words each and not with highly specialized terminology.	One compulsory text and three out of approximately eight texts, covering a variety of fields	Two texts with about 175 to 185 words each. One compulsory and general in nature. The other chosen from two others.	One text of approximately 1000 words in one of the subject areas offered. If rejects the first, a second is sent. If rejects the second, a third is sent.	<ul style="list-style-type: none"> <li>• Two out of three translations of a Non-Specialized Text tasks, each approximately 250 words</li> <li>• One Revision of a Non-Specialized Translation task, approximately 250 words</li> </ul>
<b>Duration</b>	3 hours	24 hours	3 hours	4 days	3 and half hours

**Table 4.** Source texts and durations

SATI and ITI notably differ from the other three in duration, number of source texts, and number of words. SATI and ITI allow their assesseees to translate the assigned source text in 24 hours and four days, respectively (Table 4). Just as CTTIC does, SATI also requires candidate translators to translate a compulsory source text along with three others out of eight source texts. To the best knowledge of the author, no data on the number of words in each source are provided on SATI's website. The highest number of words is observable in ITI's source text, which amounts to around 1000 words. The assessee is required to translate a single document in ITI's assessment. However, if the assessee rejects the first one, then a second is sent. If he/she rejects the second one, a third/last one is provided. If further texts are not available, the test is postponed until other texts are obtained.

As mentioned above, different practices exist in relation to the number of texts, the number of words, and the time allocated for translation tasks. In such a case, the representativeness of the source text may be discussed. CTTIC's source texts of 175 to 185 words each may raise some questions about their representativeness of the entire text if excerpted from a longer text. This might result in the exam's failure to be valid enough to reveal data on the linguistic and textual competencies of the assesseees because the source text may not contain enough linguistic challenges to activate assesseees' linguistic skills and may be too short to exhibit textual characteristics of a text in a particular field. On the other hand, a source text of 1000 words may be expected to offer an adequate number of translation challenges, but its assessment can take a considerable amount of time.

The discrepancy between the numbers of words, the numbers of source texts and the length of duration across the assessments offered by the translators organizations in question indicates that there is not a universally accepted 'best practice' in terms of these parameters.

### 7.3. Assessor, marking, and grading

**7.3.1. Assessor.** Assessment of literary vs. non-literary texts differ greatly in the evaluative approach adopted as explained above. Evaluation of a literary translation may tolerate and even favor subjectivity from a hermeneutic perspective based on the premise that every reading of a source text is unique, and translators produce their target texts accordingly. On the other hand, an objective and, if applicable, intersubjective attitude should be adopted in the assessment of non-literary translation because



translations of this kind are not supposed to be interpreted differently by each and every reader since the communicated information requires proper execution of the instructions contained in it, which should be comprehended in the same way by almost the entire intended readership. Therefore, the evaluation-related skills and knowledge of assessors as active agents of an assessment process greatly matters in conducting an objective, reliable and valid assessment.

	ATA	SATI	CTTIC	ITI	NAATI
<b>Assessor</b>	<ul style="list-style-type: none"> <li>All trained to make their grading objective and uniform.</li> <li>Two assessors.</li> <li>Third assessor in the case of disagreement between the first two.</li> <li>Fourth assessor on appeals</li> </ul>	<ul style="list-style-type: none"> <li>Double blind assessment</li> <li>Independently and anonymously by two separate assessors, generally accredited members of the Institute.</li> <li>Outsourced if necessary.</li> <li>Third assessor in the case of disagreement between the first two.</li> </ul>	<ul style="list-style-type: none"> <li>Dual marking, triple in the event of an appeal.</li> <li>Independently by assessors recruited for each language as far as possible from different provinces.</li> <li>The great majority of markers experienced certified members, accustomed to translating, revising and evaluating translation work.</li> </ul>	<ul style="list-style-type: none"> <li>Simultaneously by two assessors, fully qualified, ITI-trained members of ITI in the chosen language combination and experienced in the subject of the applicant's assessment if appropriate.</li> <li>Third assessor in the case of disagreement between the first two.</li> </ul>	<ul style="list-style-type: none"> <li>Independently marked by at least two trained assessors</li> <li>Additional assessors in the case of disagreement between the first two.</li> </ul>

**Table 5.** Qualifications of assessors

The assessors of the organizations herein can be categorized into two: trained and untrained assessors. Table 5 shows that ATA, ITI, and NAATI assign assessors that are trained in grading and marking for the assessment of target texts produced in the examination. It is important for assessors to receive assessment training to “make their grading objective and uniform”, in other words to make sure that each assessor makes the same or a similar sense out of the applicable standards and implements them accurately, which is supposed to boost the objectivity and intra- and inter-rater reliability of an assessment tool since training helps assessors pass the same or similar reasoned judgments in an assessment task instead of making impression-based subjective decisions and continuing to do so in the further evaluative practices. Training also makes assessors grow more sensitive toward translation phenomena to be considered in assessment. Since they will be trained on what a translation problem, translation solution, and translation error should be, assessors can be expected not to miss or overlook these translational indicators. Another benefit of training for assessors is the fact that they may inculcate in graders the urge to apply certain criteria by equipping them with the declarative knowledge of what to expect in a target text and the procedural knowledge of how to analyze it, which may be called assessment competence. Thanks to theoretical discussions and practical training, assessors will know when to employ their evaluative skills, will not waste time groping through the target text and the assessment process, will be able to make to-the-point decisions on the nature of a given phenomenon, and will be able to place it in the most appropriate category.

On the other hand, SATI and CTTIC assign “generally accredited members of the institute or outsourced assessors if necessary” and “experienced certified members, accustomed to translating, revising and

evaluating translation work”, respectively. SATI states that it generally employs its accredited members of the institute; however, it is dubious whether accreditation evidences an assessor’s ability to evaluate a translated text professionally. In relation to the outsourced assessors, the institute claims that it “makes sure that such persons are appropriately qualified to undertake the assessment and understand the basis on which the assessment should be done. The persons utilized would be university lecturers and reputable practitioners, for example”. Yet it presents no transparent account as to what kind of a selection process an external assessor is subjected to. Besides, being a university lecturer or a reputable practitioner may not be qualifications that guarantee proficient assessment performance either because they are not likely to be expert practitioners of the organization’s assessment guidelines or because their previous assessment experiences with other evaluative settings may interfere with their assessments at SATI. Similarly, CTTIC uses its “experienced certified members”, who are “accustomed to translating, revising and evaluating translation work”. A member’s certification in translation may not assure a reliable and valid assessment and “being accustomed” connotes incidental acquisition of a skill. “A central feature of [...] incidental learning is learning from and through experience” (Marsick and Watkins 1990, 14). It is “a byproduct of some other activity, such as task accomplishment, interpersonal interaction, sensing the organizational culture, or trial-and-error experimentation” (Marsick and Watkins 1990, 6). Unlike structured learning, incidental learning may cause an individual to miss certain critical aspects of a topic. Therefore, assessments by “accredited” and “certified” members may prove partial, which may pose problems with validity, because the assessment process may partially cover what it is supposed to assess while lacking some potentially critical translation phenomena. Hence, training assessors may contribute greatly to promoting the objectivity, reliability, and validity of an assessment process.

Table 5 shows that all five organizations assign two assessors to grade the translations of the candidates. This undoubtedly increases the interrater reliability because two assessors mean two evaluative cases and approximating their assessment skills through training is likely to result in their passing the same or similar judgments about the same translational phenomenon. They all require assessors to mark independently<sup>18</sup>, and in the event that the two graders submit contradicting scores or grades, they resort to a third one to make the final decision. The inclusion of a third assessor suggests that assessors may differ in the way they operationalize the provided criteria and hesitate or fail to employ them. Such an incident may indicate that one of the two has failed to perform a valid assessment, yielding contradictory results and requiring the inclusion of a third assessor to achieve an intersubjectively reliable result.

*7.3.2. Marking and grading.* As is clear from Table 6, each organization requires assessors to follow criteria guidelines while marking and grading. Broadly speaking, each guide governs how assessors should categorize and mark indicators of translational phenomena and assign a grade based on the marked indicators.

Table 6 shows that there are two types of indicators the organizations based their scoring on: errors and solutions. ATA, SATI, and CTTIC’s assessments are based on errors, whereas ITI builds its assessment on both errors and solutions. NAATI too operationalizes solutions and errors as assessment criteria, but while it uses errors to assess the quality of the candidates’ translations, in the revision step of the assessment it requires them to upgrade a given translation by identifying errors and suggesting solutions. It is clear from this finding that error identification is sine qua non for translation quality

18 ATA requires two graders “to consult with one another to determine the source of the discrepancy and whether it can be reconciled” [i]f the scores differ by more than 15 points, the pass/fail results are different, or the scores fall within the borderline range of 15 to 25 error points.” (Koby and Champe 2013, 168-169).

assessment because error counts in microtextual analysis are regarded “as a justification for a negative assessment” (Williams 2009, 6), which can particularly prove beneficial in the event that there is already a highest score assigned, as is the case in 100 points of CTTIC and 95 points of ITI, and error-based marks are deducted from this score. In contrast to CTTIC, ITI also marks “excellent” renderings from which assessed translators can accumulate a maximum score of 15.

One of the critical concerns over the operationalization of errors and solutions is the degree of seriousness and acceptability, respectively. ITI uses the term “excellent rendering”, which occurs when “the applicant has excelled in overcoming a particular translation issue but has not resorted to using the most straightforward renderings that the applicant would be expected to translate correctly without any difficulty”. Two issues may arise in relation to “excellent rendering”: (1) Without pre-determining translation problems solutions to which are scored, assessor’s judgment on the criticality of a translation problem would remain subjective. Besides, (2) not every translator can produce ‘fully acceptable’ target text segments. There might be cases where a solution suggested by a translator for a translation problem may not be as acceptable as the one by another translator. Thus, more than one point should be available to apply to such acceptability levels. ITI adds 1 or 2 points per “excellent rendering” to the cumulative score. The word “excellent” may be questioned in that it assigns two different points to an “excellent” translation. The same proposition holds true for error grading as well.

	ATA	SATI	CTTIC	ITI	NAATI
<b>Marking and Grading</b>	<p>Guided by a Flowchart for Error Point Decisions and a Framework for Standardized Error Marking</p> <ul style="list-style-type: none"> <li>• Error-based and “quality points”</li> <li>• Transfer/strategic errors and mechanical errors</li> <li>• Errors marked on a scale of 1, 2, 4, 8, or 16 error points and quality points up to 3.</li> <li>• 17 and fewer points: PASS</li> </ul>	<p>Guided by “Guidelines on Marking”</p> <ul style="list-style-type: none"> <li>• Error-based</li> <li>• Major and minor errors</li> <li>• Fewer than some combinations of major and/or minor errors: PASS</li> </ul>	<p>Guided by “Marker’s Guide”</p> <ul style="list-style-type: none"> <li>• Error-based</li> <li>• Major and minor mistakes</li> <li>• Translation and language errors</li> <li>• Marked out of 100</li> <li>• 70 and above: PASS</li> </ul>	<p>Governed by “a set of marking conventions and criteria” and a commentary by the assessee</p> <ul style="list-style-type: none"> <li>• Solution- and Error-based</li> <li>• Marked out of 95</li> <li>• Error: 2 points, 1 point</li> <li>• 63 and above: PASS</li> </ul>	<p>Guided by “a rubric of 5 bands across the specified skill areas”</p> <ul style="list-style-type: none"> <li>• Solution- and Error-based</li> <li>• Transfer and linguistic errors</li> <li>• Solutions only marked in “Revision” task</li> <li>• BAND 3 and above: PASS</li> </ul>

**Table 6.** Marking and grading rubrics

Assuming that there are levels to errors, each error occurring at different levels should be assigned different marks. Table 6 presents two types of errors based on category and seriousness. Category-based errors are transfer and language errors, while seriousness-based errors are major and minor errors. Therefore, one should expect to see major or minor transfer or language errors. How the organizations handle these error types notably differs. ATA categorizes the errors in the target translation of the assessee as transfer/strategic and mechanical [/language] errors and assigns 1, 2, 4, 8, or 16 error points to translation errors and 1, 2, and 4 points to linguistic errors depending on how critically the error affects the semantic level of the target text. ATA allows its assessors to “stop marking errors when score

reaches 46 error points” and “a passage with 17 or fewer points is awarded a grade of PASS<sup>19</sup>” (Koby and Champe 2013, 168). SATI counts the number of minor and major errors instead of assigning points. In SATI's assessment, “two or more major errors” or “one (or more) major error and seven or more minor errors” add up to “FAIL”. CTTIC bases its assessment on minor and major translation and linguistic errors, marks out of 100 points, and sets a PASS mark of 70. It assigns -10 points and -5 points to major and minor translation errors, respectively, and -10 points, -5 points, and -3 points to major and minor language errors (based on two seriousness levels of minor language errors). ITI employs errors and solutions to mark a final score and assesses the candidates out of 95 points. It deducts 1 or 2 points from 95 depending on the seriousness of the error and adds up 1 or 2 points in the case of an “excellent” rendering. In the end, it requires its assessees to “end up with a score of 63 or above to PASS<sup>20</sup>”. NAATI assesses the translation quality of its assessees based on “a rubric of 5 bands”. “Band 1 represents the highest level of performance and Band 5 the lowest”. It awards “PASS” to the assessee with a performance of at least Band 2 or Band 3 in four translation-related skills, i.e. Meaning Transfer Skill, Follow Translation Brief, Application of Textual Norms and Conventions along with one language skill, i.e. Language Proficiency Enabling Meaning Transfer - Target Language”.

These five organizations take into account their own assessment guides as they assess the quality of the submitted target texts. The common grading practice is to award “PASS” or “FAIL” as the final grade, but in what they differ from each other is the way they mark indicators.

	Least Critical				Most Critical
<b>ATA</b>	-1	-2	-4	-8	-16
<b>ITI</b>	-1	---	---	---	-2
<b>CTTIC</b>	-3*	---	-5	---	-10
<b>SATI</b>	Minor	---	---	---	Major
*For linguistic errors only					

**Table 7.** Shading scale of errors

Table 7 demonstrates that ATA, ITI, and CTTIC use points to differentiate between various errors types. Rather than point-based scoring, SATI enumerates the minor and major errors, while NAATI uses a rubric of 5 bands and categorizes the performances of its assessees into these five bands (not shown in Table 7 because each band do not refer to a point). ITI deducts 1 or 2 points according to the seriousness of the error. Yet its points represent the two extremities of the scale, which may suggest that its scoring system may fail to cover some of the seriousness levels of errors in between. Likewise, SATI takes minor and major errors into consideration to be able to award a pass mark. NAATI opts for employing 5 bands

19 My capitalization.  
20 My capitalization.

to categorize the errors and assesses exam takers' performances and translations' quality accordingly. In other words, it does not employ points as ATA, ITI, and CTTIC do.

One likely concern about two-point scoring, such as “(-1) and (-2)” or “(-5) and (-10)” is that even though each extremity represents the lowest and highest level achievable, (-1) and (-2) do not look as if they refer to two distant ends of a scale but two neighboring grids. There is only a one-point distance between a minor and major error, which cannot help distinguish between a minor and major error. On the other hand, a marking of (-5) and (-10) manages to demonstrate the differences in the criticality of the two error types but just assumes that error can only occur at two levels of seriousness just as the assignment of (-1) and (-2) does. Hence, regardless of how comprehensive the rubrics are and whether the assessors have received training, their assessments may turn out to be partially valid, for they will potentially miss the intermediary criticality levels. In contrast, ATA's marking system both offers points at five levels of seriousness and assigns exponentially increasing points, which allows the scoring grid to ably represent the seriousness of errors.

### **8. Conclusion and suggestions**

The present paper discussed reliability, validity, and objectivity of translation quality assessment by adopting a critical perspective on professional practices intended to assess the quality of non-literary translations. To this end, it analyzed the examinations of The American Translators Association (ATA), The South African Translators' Institute (SATI), The Canadian Translators, Terminologists and Interpreters Council (CTTIC), The Institute of Translation and Interpreting (ITI), and The National Accreditation Authority for Translators and Interpreters (NAATI). The study analyzed the assessments of these five organizations in consideration of such parameters as purpose of assessment, purpose of translation task, workplace, workbench, sources, source text, duration, assessor, marking, and grading. The results showed that these five organizations' examinations suffer from several validity-, reliability, and objectivity-related issues in varying degrees of criticality.

The analyses revealed that the organizations offer examinations for two purposes: to award certificate (ATA, SATI, CTTIC, and NAATI) or membership (ITI). It was observed that while doing so, ATA, SATI, and NAATI provide their assessees with translation briefs, but CTTIC and ITI were found to offer none. It is considered that **this may degrade the validity and reliability** of CTTIC's and ITI's examinations and also **hinder assessors from passing objective judgments**. Providing a translation brief with assessees is likely to promote examinations' validity, reliability, and objectivity by helping candidates make macro- and micro-decisions that assessors expect to see in an exam paper because assessors too are expected to take the same translation brief into account as they evaluate the papers.

It was also revealed in the present study that the organizations require assessees to translate one to four texts of varying word counts, e.g., 175 to 1000 words, in varying durations ranging from three hours to four days. This finding shows that there is not an agreed-upon text length and duration. Short texts may not contain an adequate number of challenges to call for assessees' translation knowledge and skills, which **may cause concerns about the validity of examinations** because translations of short texts may not be able to present adequate data on assessees' linguistic and textual competencies. The results showed that SATI and ITI allow their candidates to complete the tests in 24 hours and 4 days, respectively. These durations are too long and thus may not be challenging enough to reveal candidates' strategic competence, which **may degrade their assessments' validity**. To decide upon the most



appropriate text and duration length, the organizations can individually or jointly conduct research, preferably by making sure that an adequately representative sampling has been created.

The study manifested that ATA, ITI, and NAATI's examinations are marked and graded by trained assessors, while SATI and CTTIC assign their members to assess the target texts produced by assesseees. Training is likely to **increase validity and intra- and inter-rater reliability**. Without training, assessors **may pass subjective/intuitive judgments** about the quality of a translation based on their "ideal" target texts and assessment **may turn out to be partially valid** because each assessor may interpret the rubrics differently. This **may also reduce the intra-rater reliability** of SATI and CTTIC's examinations. It was discovered that all of these five organizations employ at least two assessors. This finding indicates that the examinations **are intersubjectively assessed** and therefore **may have a high level of inter-rater reliability**. It should be kept in mind that training assessors can make assessments more valid and reliable because training on how to operationalize the respective marking and grading instructions helps assessors grow more sensitive to the same or similar translational and linguistic errors, which is expected to **render assessments more objective** and **boost their inter-rater reliability** since assessors will most probably make the same or a similar decision on a specific translational phenomenon.

It was observed that these five organizations award a "PASS" or "FAIL", which are basically awarded based on transfer and language errors (as categories) and major and minor errors (as levels of criticality). What they notably differ in is "PASS" scores and points awarded to errors. ATA, CTTIC, ITI, and SATI were found to assign five, three, two, and two criticality levels to errors, respectively. NAATI was realized to employ five categories, *viz.* 'bands'. Since translation errors do not occur as 'black and white', more than two points should be assigned to errors in order to show the other seriousness levels between two extremities. The results also manifested that the distance between the scores for transfer errors in ITI, CTTIC and SATI's assessments are either too small (e.g., -1 and -2) or too far (e.g., -5 or -10). In such a scoring system, assessors may feel forced to choose between two points because they are not provided with intermediary points. In this case, even if they are supplied with well-organized comprehensive rubrics and have received training on how to operationalize the instructions therein, their assessments **may still be partially valid and reliable**. To sort out these concerns, a scoring able to foreground a vast range of error criticality and allowing for a better discernment between various levels of errors should be adopted. In this sense, ATA's scoring can be presented as an applicable method of scoring. Another solution to marking and scoring differences can be active communication and interaction among the organizations. They may meet and negotiate or carry out research projects either individually or collectively not only to come up with the most viable scoring system but also to produce the most objective, valid and reliable assessment as applicable to their respective settings as possible.

Based on the findings of the present study, an assessment should be conducted by trained assessors who can ably follow pre-determined instructions and criteria and adopt a macro- and micro-analytical perspective when studying a target text. It should also be objective and inter-subjective, thus allocentric, which potentially translates into a reliable and valid assessment tool. To save on time and labor and "as a justification for a negative assessment", they can be built on errors, yet marking and grading should be arranged in a way to represent a diverse range of error criticality and to allow for a distinct differentiation between two neighboring grids.

The current paper analyzing the translation quality assessment guidelines and rubrics of five translators organizations is considered to make considerable contributes to the efforts to produce valid, reliable,

and objective translation quality assessment tools and likely to motivate translators organizations, regardless of whether mentioned in this study, to revise their assessment practices and to collaborate with each other to produce examinations consisting of shared rubrics and criteria by eliminating major differences.

This paper focuses on translation quality assessments and the assessment procedures adopted by five “major” translators organizations. Future research can be concentrated on revealing what members, assessors, and directors of translators associations think about the suggestions available in this research study. It would be interesting to ask the assessors of the five organizations here to comment on each other’s examination processes. Moreover, it may prove fruitful to investigate which set of rubrics translation scholars would consider the most applicable, reliable, valid, and objective. A similar study can be carried out on the rubrics and guidelines of interpreters associations.

### References

- “A Guide to the ATA Certification Program,” American Translator Association, accessed October 4, 2018, [http://atanet.org/certification/aboutcert\\_overview.php](http://atanet.org/certification/aboutcert_overview.php).
- “Accreditation Examinations: Guidelines on Marking,” South African Translators’ Institute, accessed October 4, 2018, [http://translators.org.za/sati\\_cms/downloads/dynamic/majors\\_minors\\_english.pdf](http://translators.org.za/sati_cms/downloads/dynamic/majors_minors_english.pdf).
- Al-Qinai, J. (2000). “Translation Quality Assessment. Strategies, Parametres and Procedures.” *Meta: journal des traducteurs /Meta:Translators’ Journal*, 45(3): 497-519. <https://doi.org/10.7202/001878ar>
- Angelelli, C. (2009). “Using a Rubric to Assess Translation Ability: Defining the Construct.” In *Testing and Assessment in Translation and Interpreting Studies: A Call for Dialogue between Research and Practice*, ed. by Claudia V. Angelelli and Holly E. Jacobson, 13-48. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Angelelli, C. (2012). “Testing and Assessment in Translation and Interpreting Studies.” In *Handbook of Translation Studies*, ed. by Yves Gambier and Luc van Doorslaer, 172-177. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- “Applicant Handbook: Assessment Guide for Translators,” Institute of Translation and Interpreting, accessed October 4, 2018, <https://www.iti.org.uk/attachments/article/992/ITI%20Applicant%20Handbook%20v4.pdf>.
- Ary, D; Lucy C. Jacobs ; Sorensen, C. (2010). *Introduction to Research in Education*. California: Wadsworth.
- “ATA Certification Program Framework for Standardized Error Marking - Version 2017,” Last updated January 27, 2017, American Translator Association, accessed October 4, 2018, [https://www.atanet.org/certification/Framework\\_2017.pdf](https://www.atanet.org/certification/Framework_2017.pdf).
- Bittner, H. (2011). “The Quality of Translation in Subtitling.” *Trans-kom*, 4 (1): 76-87. [http://www.trans-kom.eu/bdo4nr01/trans-kom\\_04\\_01\\_04\\_Bittner\\_Quality.20110614.pdf](http://www.trans-kom.eu/bdo4nr01/trans-kom_04_01_04_Bittner_Quality.20110614.pdf).
- Bowker, L. (2000). “A Corpus-based Approach to Evaluating Student Translations.” *The Translator* 6(2): 183-210. <https://doi.org/10.1080/13556509.2000.10799065>.
- Castilho, S. ; Doherty, s. ; Gaspari, F. ; Moorkens, J. (2018). “Approaches to Human and Machine Translation Quality Assessment.” In *Translation Quality Assessment: From Principles to Practice*, ed. by Joss Moorkens, Sheila Castilho, Federico Gaspari, and Stephen Doherty, 9-38. Cham: Springer International Publishing.
- “Certified Translator Test: Instructions for Candidates,” Last updated October 4, 2018, National Accreditation Authority for Translators and Interpreters, accessed October 5, 2018, <https://www.naati.com.au/media/1951/ct-candidate-instructions-finalpdf.pdf>.

- Cohen, L. ; Manion, L. ; Morrison, K. (2000). *Research Methods in Education*. London: Routledge.
- “Computerized Certification Exam,” American Translator Association, accessed October 4, 2018, [https://atanet.org/certification/aboutexams\\_computerized.php](https://atanet.org/certification/aboutexams_computerized.php).
- “CTTIC Candidate’s Guide for the CTTIC Standard Certification Examination in Translation,” Last updated March, 2005, Canadian Translators, Terminologists and Interpreters Council, accessed October 4, 2018, [http://www.cttic.org/examDocs/CandidateGuide\\_En1011.pdf](http://www.cttic.org/examDocs/CandidateGuide_En1011.pdf)
- “CTTIC Standard Certification Translation Examination: Marker’s Guide,” Last updated March, 2005, Canadian Translators, Terminologists and Interpreters Council, accessed October 4, 2018, <http://www.cttic.org/examDocs/guide.markersE.pdf>.
- English Oxford Living Dictionaries, s.v. “quality,” accessed October 07, 2018, <https://en.oxforddictionaries.com/definition/quality>.
- “Flowchart for Error Point Decisions - Version 2009,” American Translator Association, accessed October 21, 2018, [http://atanet.org/certification/aboutexams\\_flowchart.pdf](http://atanet.org/certification/aboutexams_flowchart.pdf).
- “Framework for Marking of Translation Accreditation Examinations,” South African Translators’ Institute, accessed October 4, 2018, [http://translators.org.za/sati\\_cms/downloads/dynamic/marking\\_guidelines\\_translation\\_english.pdf](http://translators.org.za/sati_cms/downloads/dynamic/marking_guidelines_translation_english.pdf)
- Hansen, G. (2003). “Controlling the Process: Theoretical and Methodological Reflections on Research into Translation Processes.” In *Triangulating Translation: Perspectives in Process Oriented Research*, ed. by Fabio Alves, 25-42. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hansen, G. (2008). “The Speck in Your Brother’s Eye – The Beam in Your Own: Quality Management in Translation and Revision.” In *Efforts and Models in Interpreting and Translation Research*, ed. by Gyde Hansen, Andrew Chesterman, and Heidrun Gerzymisch-Arbogast, 255-280. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hatim, B.; Mason, I. (2005). *The Translator as Communicator*. London/New York: Routledge.
- Holmes, J. S. (2000). “The Name and Nature of Translation Studies.” In *Translation Studies Reader*, ed. by Lawrence Venuti, 172-185. London/New York: Routledge.
- House, J. (1997). *Translation Quality Assessment: A Model Revisited*. Tübingen: Gunter Narr Verlag.
- House, J. 2001. "Translation Quality Assessment: Linguistic Description versus Social Evaluation." *Meta: Journal des Traducteurs/Meta: Translators' Journal* 46 (2): 243-257. doi:10.7202/003141ar.
- House, J. 2009. “Quality.” In *Routledge Encyclopedia of Translation Studies*, ed. by Mona Baker and Gabriela Saldanha, 222–225. London/New York: Routledge.
- “Information Regarding the Accreditation Examinations,” South African Translators’ Institute, accessed October 4, 2018, [http://translators.org.za/sati\\_cms/downloads/dynamic/sati\\_accreditation\\_general\\_guidelines\\_english.pdf](http://translators.org.za/sati_cms/downloads/dynamic/sati_accreditation_general_guidelines_english.pdf).
- Koby, G. S. ; Champe, G. G. (2013). "Welcome to the Real World: Professional-Level Translator Certification." *The International Journal for Translation and Interpreting Research* 5(1): 156-173. doi:ti.105201.2013.a09.
- Koby, G. S. ; Fields, P. ; Daryl R. Hague, Arle Lommel, and Alan Melby. 2014. “Defining Translation Quality.” *Revista Tradumàtica: Tecnologies de la Traducció* 12: 413-420. [https://ddd.uab.cat/pub/tradumatica/tradumatica\\_a2014n12/tradumatica\\_a2014n12p413.pdf](https://ddd.uab.cat/pub/tradumatica/tradumatica_a2014n12/tradumatica_a2014n12p413.pdf).
- Krein-Kutiile, M. (2003) “Equivalence in Scientific and Technical Translation: A Text-in-Context-based Study.” PhD Diss., European Studies Research Institute, University of Salford.

- Lommel, A. ; Popović, M. ; Burchardt, A. (2014). "Assessing Inter-Annotator Agreement for Translation Error Annotation." In Proceedings of MTE Workshop on Automatic and Manual Metrics for Operational Translation Evaluation, LREC 2014, Reykjavik, Iceland.
- Marsick, V. J. ; Watkins, K. E. (1990). Informal and Incidental Learning in the Workplace. London/New York: Routledge
- Martínez-Melis, N., Hurtado Albir, A. (2001). "Assessment in Translation Studies: Research Needs. Meta:Journal des Traducteurs/Meta: Translators' Journal 46(2): 272-287. <https://doi.org/10.7202/003624ar>.
- Morin, K. H. ; Barbin, F. ; Moreau, F. ; Toudic, D. ; Phuez-Favris, G. (2017). "Translation technology and learner performance: Professionally-oriented translation quality assessment with three translation technologies". In Translation in Transition: Between Cognition, Computing and Technology, ed. by Arnt Lykke Jakobsen and Bartolomé Mesa-Lao, 207-233. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- "Policy for the Use of Equipment and Reference Materials in Certified Translator Tests and Certified Advanced Translator Tests," Last updated March 8, 2018, National Accreditation Authority for Translators and Interpreters, accessed October 5, 2018, <https://www.naati.com.au/media/1952/policy-use-equipment-reference-materials-trans-testspdf.pdf>.
- "Procedure to Appeal Certification Exam Review," Last updated November 2009, American Translator Association, accessed February 12, 2019, [http://atanet.org/certification/aboutreview\\_appeal.php](http://atanet.org/certification/aboutreview_appeal.php)
- Rothe-Neves, R. (2002) . "Translation Quality Assessment for Research Purposes: An Empirical Approach." *Cadernos De Tradução* 2(10): 113-131.<https://doi.org/10.5007/%25x>.
- Schäffner, C. (1997). "From 'Good' to 'Functionally Appropriate': Assessing Translation Quality." *Current Issues in Language and Society* 4(1). <https://doi.org/10.1080/13520529709615476>.
- Schrijver, I. ; Van Vaerenbergh, L. (2012). "An Exploratory Study of Transediting in Students' Translation Processes". *HERMES - Journal of Language and Communication in Business* 25(49): 99-117. <https://doi.org/10.7146/hjlc.v25i49.97740>.
- Thelen, M. (2008). "Translation Quality Assessment or Quality Management and Quality Control of Translation?." In Translation and Meaning – Part 8, ed. by Barbara Lewandowska-Tomaszczyk and Marcel Thelen, 411- 424. Maastricht: Hogeschool Zuyd.
- Triandis, H. C. (1983). Allocentric vs. idiocentric social behavior: A major cultural difference between Hispanics and the mainstream. Technical Report No. 16. Department of Psychology, Urbana-Champaign: University of Illinois.
- Williams, M. (1989). "The Assessment of Professional Translation Quality: Creating Credibility out of Chaos." *TTR: Traduction, Terminologie, Rédaction* 2(2): 13-33. <https://doi.org/10.7202/037044ar>.
- Williams, M. (2009). "Translation Quality Assessment." *Mutatis Mutandis* 2(1): 3-23. <http://aprendeenlinea.udea.edu.co/revistas/index.php/mutatismutandis/article/view/1825/1609>.
- Yazıcı, M. (2007). *Yazılı Çeviri Edinci*. İstanbul: Multilingual
- Zehnalová, J. (2013). "Tradition and Trends in Translation Quality Assessment". In Tradition and Trends in Trans-Language Communication, ed. by Jitka Zehnalová, Ondřej Molnár, and Michal Kubánek, 41-58. Olomouc: Univerzita Palackého.