Translation Quality Assessment Rubric: A Rasch Model-based Validation

Aynaz Samir[1*], Mona Tabatabaee-Yazdi[1]

Abstract

The present study aimed to examine and validate a rubric for translation quality assessment using Rasch analysis. To this end, the researchers interviewed 20 expert translation instructors to identify the factors they consider important for assessing the quality of students' translation. Based on the specific commonalities found throughout the interviews, a 23-item assessment rubric was constructed on a four-point Likert scale. More specifically, this study used the Rasch rating scale model for polytomous data to investigate the psychometric properties of the rating scale in terms of dimensionality, reliability, use of response category, and sample appropriateness. Then, a translation exam was administered to 60 translation students at the BA level in Iranian universities. Following this, the rubric was employed to assess the quality of students' translation. The results revealed that the Rasch model fits the data well. Thus, the findings of the study indicated that the rubric is potentially valid and useful, and can be used as a measure of translation quality assessment in the Iranian context.

Keywords: Polytomous Data, Rasch Model, Rubric, Translation Quality Assessment, Validity

1. Introduction

In any educational setting assessment is a process of gathering data regarding teaching and learning to make decisions about how to improve students' learning outcomes and the strategic objectives of the program, and remove program shortcomings (Genesee 2002; Genesee & Upshur, 1996; Rea-Dickins & Germaine, 1993). Although currently, an essential section of the translator training process is assessment (Stobart & Gipps, 1997), few empirical studies have been carried out on assessing the processes and products of translation (Angelelli & Jacobson, 2009). Instructors also devote a minimum portion of their time and attention to activities related to assessment (Bachman, 2014). Sharififar, Beh-Afarin, and Younesi (2018) also stated "the criteria considered to assess students' translations are teacher-made because most of the lecturers chose them based on their experience in this field and not based on their knowledge"

[1] English department, Tabaran Institute of Higher Education, Mashhad, Iran.
* (corresponding author: aynazsamir@yahoo.com)

(p. 52). As rightly declared by Arango-Keeth and Koby (2003) translation quality assessment (TQA) in the field of Translation Studies is still the least developed. There is a lack of research on assessing the quality of translation in the field of research and pedagogy (Angelelli & Jacobson, 2009). Conde (2012, p. 68) also stated that "evaluation is still a field in which much remains to be explored". Hatim and Mason (1997, p. 197), said that "the assessment of translator performance is an activity which, despite being widespread, is under-researched and under-discussed". Even though most of the previous theoretical researches on translation quality assessment concentrated on objective assessment (House, 1997; Newmark, 1988; Wilss, 1998), it is difficult to conduct a reliable and effective objective assessment (Khorami & Modarresi, 2019). The major reason could be the lack of direct observation and explanation of personal, social, and discoursal aspects of translation (Beeby, 2000). Besides, since there has been a gap among translation theory and translation practice (Snell-Hornby et al., 1994), "scholarly approaches to translation evaluation have not yet been able to provide help for practical quality assessment because they do not account for the reality of translating and translations" (Lauscher, 2000, p. 158). As truly stated by Hanifehzadeh and Farahzad (2016) "finding the procedures for reducing the observed subjectivity or even objectivity of constructed scales can be an important concern for interested scholars in the field" (p. 84). Hence, it seems that more relevant research studies are likely to be done to bridge the gap between translation practice and translation theory.

Moreover, taking account of translation assessment, rubrics provide more holistic, analytic, and systematic scoring (Angelelli, 2009). Rubrics give "descriptive statements of behaviors that candidates may exhibit in a particular sub-component" (Angelelli, 2009, p. 39). An assessment rubric has the whole sub-components which make the main constructs. Since rubrics are used to assess nearly all products or performance holistically (Moss & Holder, 1988; Walvood & Anderson, 1998), it plausible to discuss its practicality for assessing translation. Moreover, in language testing, assessment rubrics are used to evaluate the main competencies in the production of language (Cohen, 1994). Considering translation assessment, competencies are specific skills or traits which are being assessed (Bachman, 2014). Indeed, the competencies in developing a systematic and holistic rubric for translation assessment should be defined and explained (Bachman & Palmer, 1996). Furthermore, a rating scale is required for measuring each of these competencies (Mertler, 2001). In this case, rubrics recognize what translation competence is being evaluated (Walvood & Anderson, 1998). Thus, by developing an assessment rubric, examiners can assess all the competencies that are related to a test (Wiggins, 1998).

However, it seems that in the field of Translation Studies few rubrics have been developed for assessing translation. In addition, few empirical research has been conducted on the assessment of translation and on the issues of measuring translation by employing reliable and valid rubric. As clearly stated by Campbell and Hale (2003), in translation assessment the issues of reliability and validity of measurements require further explanations. Furthermore, as Muñoz (2012, p. 170) declared, "We gained much insight into translators' mental life, but there has been very little construct-validating research". Knowing how graders make the decision and creating operational test tools depend on valid test constructs that are vital for real exam

organization and research projects on the issues regarding the translation assessment. Therefore, a valid and effective assessment rubric for evaluating the quality of translation is essential to a successful education (Qomaria & Thahara, 2015).

Consequently, the present study aimed to develop an educational rubric for assessing the quality of translation and investigate the extent to which the diagnostic information of the rubric, was a reliable, accurate, and discriminant method of assessing translation performance. To do so, the educational rubric for translation assessment was developed based on the perspective of expert translation instructors and reviewing of the literature focusing on translation quality assessment. Besides, the researchers analyzed the actual application of the rubric in translation courses at the BA level in Iranian universities to validate a 23-item assessment rubric, utilizing the Rasch measurement model which is usually applied for analyzing measurement instruments data and examining tests and constructs validity in social sciences (Baghaei, 2008).

2. Review of the Literature

2.1 Translation Quality

Fundamentally, to progress in the translation industry, the quality of translation should meet the defined criteria. These criteria are determined by the stakeholders including translation theorists, providers, producers, end-users, or the requesters of the services (Techno Translation Studio, 2016). Translation quality is generally a problematic concept in translation research (Qomaria & Thahara, 2015). On defining Translation Quality (TQ) Koby et al. (2014) introduced two contrasting definitions of TQ, labeled as "Broad" and "Narrow". According to them the broad view of TQ supports this component of the definition "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" (p. 416). Additionally, Koby et al. (2014) believe the narrow view of TQ has the following definition:

> A high-quality translation 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. (p. 417)

In this regard, the quality of translation can be judged by comparing ST and TT to the criteria determined by the stakeholders (Techno Translation Studio, 2016). However, some scholars stated that the broad TQ definition is sufficient and some others believed the narrow view of TQ is acceptable by the stakeholders of the translation industry (Koby et al., 2014). Inclusively, the quality of translation is revealed in some ways for instance "a translation must be idiomatically, grammatically and terminologically correct" (Ørsted, 2001, p. 445). "The quality of translation is also determined by the accuracy of the delivery of messages from the source language to the target language" (Akhiroh, 2013, pp. 42-43). According to Akbari and Shahnazari (2015), the quality of translation determines by the translator's knowledge

regarding source language (SL) and target language (TL), the intention of the reader, and context (p. 445). Additionally, Mahmoudian et al. (2011, p. 95) defined the quality of translation as "the suitability of words and sentences from aspects of clarity, use of common language, and similarity of content/meaning".

2.1.1 Translation Quality Assessment

In the field of Translation Studies assessing the quality of translation is seriously important. As declared by House (2015) "Translation quality assessment can thus be said to be at the heart of any theory of translation" (p. 1). However, research into TQA remains one of the most difficult phases of investigation in Translation Studies since it is shown that translation scholars have some difficulties in assessing a translated text (Modarresi & Ghoreyshi, 2018; Williams 2009). "The main problem seems to reside in how to express quality or what measure should be used for the quality of a translation" (Gharacheh, 2005, p. 20). In this regard, during the past decades, some translation theorists introduced some approaches to assess the quality of translation. Thus, based on those assessment approaches, some assessment models and rubrics were developed. One of these translation quality assessment models was developed by Broeck (1978) paid attention to literary translation. He tried to establish the same functional equivalence among literary source text (ST) and target text (TT). Kim (2009) as well presented a TQA model known as Meaning-Oriented Assessment of Translations to quantitatively assess the quality of the translated texts at the university level. Moreover, House (1997, 2015) developed a TQA model named Translation Quality Assessment. She (2006, p. 325) believed that "Translation can be defined as the replacement of a text in a source language by a semantically and pragmatically equivalent text in a target language. An adequate translation is thus a pragmatically and semantically equivalent one." Hence, House (1997) developed a TQA model by focusing on the pragmatic theories (as cited in Baker & Saldanha, 2009). She introduced this TQA model based on comparative analysis of ST and TT, and the analysis of the linguistic-situational features of the two texts (Munday, 2001). In her TQA model, the main requirement for equivalence of ST and TT is that the translated text must have an ideational and interpersonal function that is corresponding to the function of the source text. The comparison of ST and TT leads to assessing the quality of TT via focusing on errors. However, to achieve the ST function, the TT must make use of the equivalent pragmatic means.

2.2 Rubric in Translation Assessment

The major step in assessing the translation is to create a quality model and after that change it into a series of metrics that assess all components of that quality. As declared by Stevens and Levi (2004, p. 3) "At its most basic, a rubric is a scoring tool that lays out the specific expectations for an assignment. Rubrics provide detailed descriptions for what constitutes acceptable and unacceptable levels of performances." Riazi (2003) stated that rubrics help translation students and teachers to recognize the assessment criteria which are unbiased and objective. In addition, rubrics provide ground for reflection, peer review, and self-assessment (Riazi, 2003). Rubrics are created by recognizing the features of translation competence, the crucial characteristics of the performance and product such as strategic competence, pragmatic

competence, textual competence, micro- linguistic competence, and following these criteria used to categorize different performance by focusing on any of such subcomponents. Muzii (2007) maintained that a series of systematic criteria evaluate the quality of translated text from various viewpoints during the process of production. Making a single metric would not disclose all problems. Developing multiple metrics that evaluate each aspect of the translated text, could help rearrange the whole framework and give a hint of which parts do not work well and which part does. Therefore, a valid and reliable rubric for assessing the quality of translation is needed to address the aforesaid issues. It is worth mentioning that, so far different rubrics were developed in the realm of Translation Studies by the translation scholars including Farahzad (1992), Beeby (2000), Waddington (2001), and Goff-kfouri (2004). For instance, Sager (1983) presented a rubric in five levels of scoring based on pragmatic error, linguistic error, and semantic error. In Mason's rubric (1997, as cited in Khanmohammad & Osanloo, 2009) four levels of errors were designed for each sentence. Later some rubrics created with an instructional framework that focused on errors and positive points. In this regard, one rubric was presented by Rico-Perez (2002) to find three levels of seriousness and six types of errors.

2.2.1 Frahzad's Rubric

In Farahzad's (1992) rubric two features, Accuracy, and Appropriateness are being checked to score each sentence (Table 1). It worth mentioning that, the unit of translation in Frahzad's (1992) rubric is a sentence. Nevertheless, accuracy in her rubric means that all information in the ST should be conveyed precisely in the target text. Appropriateness also means that each sentence structure must be correct, and sound native and fluent. Half of the score is given to unnatural translation that conveys the meaning of source text, whilst no score is given to inaccurate translations. In addition, error recognition receives one score and another one score is given for error correction. Two different scoring ways were introduced by Farahzad (1992) for a long text. The two methods of scoring are as following:

A. In the first method of scoring, the TT is scored holistically. In this regard, the whole text is the unit of translation, and the grader allots score to each significant factor. For example, (a) accuracy receives 20 percent, (b) appropriateness receives 20 percent, (c) naturalness receives 20 percent, (d) cohesion receives 20 percent, and (e) style of discourse/choice of words receives 20 percent of the marks to be awarded.

B. In the second method of scoring, the examiner read the TT two times. Firstly, the examiner checks the accuracy and appropriateness of the TT. The unit of translation in this method is sentence and clause. The marker of a sentence is a verb. Therefore, each verb marks a score in the ST. When the ST has one main clause and a subordinate clause, one score is given for the main clause and another score for the subordinate clause. Secondly, the examiner checks the cohesion (e.g. appropriate use of pronouns, linkages, etc.) and style (e.g. choice of words, grammatical structures, etc.) in each sentence and clause of the TT. However, no score is given to a sentence that does not convey the content. Moreover, no score is given to a sentence in which its structure distorts meaning. The translation gives half of a score if it conveys the message, but in an unnatural grammatically structure. Nevertheless, Farahzad (1992) declared that the examiner cannot precisely check and score the cohesion and style at

the sentence level or clause level. Since the components of cohesion and style of discourse are used all over the text.

Table 1

Farahzad's (1992) model based on sentence and clause as the unit of translation

	Accuracy and Appropriateness		Cohesion and Style			
Sentences	Main Clause	Sub-Clause	Appropriate use of pronouns	Linkages	Choice of Words	Grammatical structures
1						
2						
3						
…						

2.2.2 Goff-Kfouri's Rubric

Goff-Kfouri (2005) stated that the examiner can choose from three options for assessing a translation including general impression, error count, and analytical grid. (a) General Impression: Even though some professional examiner is capable enough to distinguish the differences between a translation which is a 62/100 and the one is 67, the general impression score is not useful for translation trainees. Since the reasons for the missing scores were not provided. (b) Error count: This method of counting errors is not recommended for assessing the TT because it does not consider the seriousness of mistakes and seldom gives a score for content. (c) Analytical grid: It was first introduced by Heaton (1990) for language courses. Then, the analytical grid was adopted for correcting the translations. It provides the examiners some criteria for correcting translation according to simple arithmetic. The TT is scored over 23 and the scores for four criteria are 1 to 5. But, the maximum score for mechanics is 3 which is weighted less than other criteria. The correction criteria are shown in Table 2.

Table 2
Goff-Kfouri's rubric

Correction Criteria	5	4	3	2	1
Fluency					
Grammar					
Terminology					
General Content					
Mechanics			X	X	

2.2.3 Waddington's Rubric

Waddington (2001) stated that different texts should be assessed differently. Therefore, Waddington (2001) in his model of TQA proposed four assessment methods. Method A and Method B are based on error analysis. Probable mistakes in Method A are classified into three headings. (a) Improper translations that impact the understanding of the ST. They are categorized into eight criteria including addition, loss of meaning, omission, nonsense, faux sens, inappropriate linguistic variation, contresens (mistranslation), and unresolved extralinguistic references. (b) Improper translating, that impact expression in TT. They are categorized into five criteria such as lexical items, grammar, spelling, text, and style. (c) Inadequate translating that impacts the transmission of primary and secondary functions of the ST. There is a distinction in these categories between serious error (-2 scores) and minor errors (-1 score).

Method B was developed to consider the negative impact of errors on the quality of the whole TT. The examiners should decide each error is a translation mistake or language mistake. If an error influences the rendering of meaning from ST to the TT, it is a translation error (subtracted 2 points). If the error does not influence, it is a language error (subtracted 1 point). Besides, the examiner should take into account the importance of the negative effect of each error on the translation.

Method C was designed as a holistic assessment method. In this method, the translation competence was considered as a whole. The examiner should take into account the three aspects of the translator's performance. Waddington (2001) designed five levels of performance in this method. Then, he determined two possible scores for each level. In this case, if a translation fully fulfills the requirements of a specific level, it receives a higher score. On the contrary, if a translation is placed between two levels but is closer to the upper level, it receives the lower score (p. 315). Waddington (2001) asserted that Method D is "a method which consists of combining error analysis Method B and holistic Method C in an appropriation of 70/30; that is to say that Method B accounts for 70% of the total result and Method C for the remaining 30%" (p. 315).

2.2.4 Beeby's Rubric

In Beeby's (2000) rubric the translation exam was scored out of 20 points. Ten scores are given for 10 special translation problems which have been found in the text. In addition, language receives 10 scores. In this regard, 1 score is subtracted for incorrect word order, tense, agreement, and syntax. Likewise, half of the score is subtracted for incorrect spelling, prepositions, and articles. Nevertheless, Beeby (2000) asserted that in this rubric the 10 scores were chosen for assessing translation competence (Table 3). In this regard, score 1 was selected for translating the headline. Score 2 was selected for typographical differences. Score 3 to 6 for transfer competence, discourse competence, and knowledge of syntactic differences among the ST and TT. Score 7 for relevance, score 8 for lexical errors; score 9 for cultural transfer; and score 10 for extralinguistic knowledge.

Table 3
Beeby's competence-based rubric

Translation Elements	Headlines	Typographolo	Transfer competence	Discourse competence	Syntactic differences	Relevance	Lexical errors	Cultural transfer	Extra-linouistic linouistic
Scores									

2.3 Translation Competence

The definition of the term competence is still being developed, with a variety of definitions in circulation. In almost all cases, scholars tended to break translation competence down into a set of interrelated sub-competencies, which can be studied in isolation (Schaffner & Adab, 2000). Therefore, translation competence is "bilingual competence, bicultural competence, creative competence, thinking competence, expressive competence, extralinguistic competence, and transfer competence" (Lei, 2006, p. 61). Nord (1999) believed that translation competence is the "translational text competence," consisting of meta-competence, text-production competence, text-analytical competence, and contrastive text competence. In addition, Neubert (2000, p. 6) stated that translation competence includes "language competence, textual competence, subject competence, cultural competence, and transfer competence." According to Kelly (2008) translation competence is a macro-competence which includes the whole attitudes, skills, abilities, and knowledge of specialized translators, who take part in translation. According to Kelly (2005), translation macro-competence can be further subdivided into seven sub-competences. Kelly (2008) has identified the following competence for translator teaching: (a) communicative and textual competence in at least two languages and cultures; (b) cultural and intercultural competence; (c) subject-matter competence; (d) professional and instrumental competence; (e) attitudinal (or psycho-physiological) competence; (f) interpersonal competence; (g) strategic competence (pp. 38-39). In Mackenzie's view (2004), translation competence includes linguistic-cultural skills, interpersonal skills, IT skills, marketing ability, even management skills, since quality in translation needs management of the whole process. Accordingly, the present study aimed to design and validate a translation quality assessment rubric to be used in the Iranian educational context.

3. Methodology

3.1 Participants

The first group of participants was translation instructors. In the first phase of the study, a semi-structured interview was conducted to discover the perspectives of 12 expert translation instructors in the Iranian context regarding the significant criteria which are required for quality translation assessment. Therefore, relying on the literature review and consultation with two experts an interview was developed including five questions such as "*what are the most and*

the least important criteria for assessing the quality of translation?", and *"what translation quality assessment criteria receive a higher score?".* The 12 expert translation instructors were selected based on four main criteria. They should have (a) translation teaching experience at universities, (b) Ph.D. degree in different fields of study in English, (c) experience in translation and translation assessment, and (d) their own published textbooks, novels, or articles in the field of translation studies focusing on translation assessment. Therefore**,** the participants were chosen by purposive sampling technique. The interviews were held in a semi-structured form and performed in a conversational style. During this phase, 12 translation instructors comprising 3 females and 9 males, from different cities of Iran, were participated (Table 4).

Table 4

Demographic profile of respondents (Instructors, N=12)

Participants	Category	Frequency	Percentage
Gender	Female	3	25%
	Male	9	75%
Educational Status	PhD	12	100%
Translation Teaching	5-10 Years	5	41.7%
	More than 10 years	7	58.3%
Field of Study	TEFL	4	33.3%
	English Translation Studies	6	50%
	Linguistics	2	16.7%
City	Mashhad	9	75%
	Torbateh-Heidarieh	1	8.3%
	Quchan	2	16.7%

Consequently, as an integral step of developing a new rubric, 60 BA students, from different major universities in Iran, were invited to translate two unseen texts from English into Persian (Table 5). They were males (40%) and females (60%) with more than three years of translation experience. The participants were selected from both State and Non-State universities. In the curriculum of English Translation at the BA level, most of the technical translation courses are presented in the sixth, seventh and eighth semesters. Therefore, in this phase of the study, 60 translation students in sixth, seventh, or eighth semesters were selected.

Table 5
Demographic profile of respondents (Students, N=60)

Participants	Category	Frequency	
Gender	Female	36	60%
	Male	24	40%
Educational Status	BA	60	100%
Translation Experience	Below 5 Years	60	100%
Field of Study	English Translation	60	100%
City	Mashhad	27	45%
	Tehran	16	26.7%
	Others*	17	28.3%

* Birjand, Bojnord, Isfahan, Kerman, Quchan, Qazvin, Shiraz, Torbat-e-Heidarieh, Yazd

3.2 Instrumentations

To achieve the aim of the present study four instruments were found valuable to collect the necessary data. The first instrument used was an interview developed by reviewing related literature on translation quality assessment methods and models. Therefore, a face-to-face interview was conducted to identify the perspective of 12 translation instructors regarding the criteria involved in assessing translation. The content validity of the interview questions was ensured by consulting two experts from the field. The interview was held in a semi-structured form and performed in a conversational style.

The translation assessment checklist was the second instrument to identify the major criteria in assessing students' translation. Hence, the thematic analysis of interview data showed that 32 assessment criteria were mentioned by all instructors. Besides, based on reviewing the related literature on the pre-established translation quality assessment rubrics, 28 important assessment criteria were chosen for assessing the quality of translation including grammar (Huot, 1990; Beeby, 2000; Wadington, 2001; Goff-Kfouri, 2005; Doyle, 2003; Hansen, 2010; Conde, 2013; Lee & Ronowick, 2014; Dewi, 2015), usage (Doyle, 2003; Khanmohammad & Osanloo, 2009; Dewi, 2015), addition (Hurtado,1995, as cited in Waddington, 2001; Vinay & Darbelnet, 1995; Williams, 2001; Doyle, 2003; Khanmohammad & Osanloo, 2009; Waddington, 2001; Angelone, 2013; Dweik & Suleiman, 2013; Lee & Ronowick, 2014; Dewi, 2015), omission (Hurtado,1995, as cited in Waddington, 2001; Vinay & Darbelnet , 1995; Williams, 2001; Doyle, 2003; Khanmohammad & Osanloo, 2009; Waddington, 2001; Angelone, 2013; Dweik & Suleiman, 2013; Lee & Ronowick, 2014; Dewi, 2015), completeness (Waddington, 2001; Doyle, 2003; Angelone, 2013; Conde, 2013; Lee & Ronowick, 2014), punctuation (Huot, 1990; Williams, 2001; Doyle, 2003; Conde, 2013; Lee & Ronowick, 2014; Dewi, 2015), terminology (Doyle, 2003; Goff-Kfouri, 2005; Angelone, 2013; Lee & Ronowick, 2014; Dewi, 2015), false friend (Hurtado,1995, as cited in Waddington, 2001; Waddington, 2001; Doyle, 2003; Angelone, 2013); spelling (Hurtado,1995, as cited in Waddington, 2001; Beeby, 2000; Waddington, 2001; Doyle, 2003;

Lee & Ronowick, 2014; Dewi, 2015), capitalization (Doyle, 2003; Dewi, 2015), faithfulness (Benjamin, 1972, as cited in Dewi, 2015; Doyle, 2003; Nord, 2005; Lee & Ronowick, 2014; Dewi, 2015), register (Halliday & Hasan, 2014; Halliday, 1978; Hurtado,1995, as cited in Waddington, 2001; Waddington, 2001; Doyle, 2003; Hansen, 2010; Khanmohammad & Osanloo, 2009; Angelone, 2013; House, 2015), genre (Hansen, 2010; Khanmohammad & Osanloo, 2009), style (Hurtado,1995, as cited in Waddington, 2001; Waddington, 2001; Williams, 2001; Doyle, 2003; Farahzad, 1992; Khanmohammad & Osanloo, 2009;), text type (Reiss, 2014; Koller, 1979), coherence (Koller, 1979; Halliday & Hasan, 2014; Doyle, 2003; Hansen, 2010; Conde, 2013;), cohesion (Halliday & Hasan, 2014; Doyle, 2003; Farahzad, 1992; Hansen, 2010; Lee & Ronowick, 2014), consistency (Doyle, 2003; Khanmohammad & Osanloo, 2009; Angelone, 2013), accuracy (Farahzad, 1992; Stansfield et al., 1992; Waddington, 2001; Doyle, 2003; Khanmohammad & Osanloo, 2009; Polliastri & Paulina, 2009), Naturalness (Farahzad, 1992, Doyle, 2003), ambiguity (Williams, 2001; Doyle, 2003; Conde, 2013; Lee & Ronowick, 2014; Dewi, 2015), fluency/readability (Farahzad, 1992; Conde, 2011; Dewi, 2015), mistranslation (Doyle, 2003; Angelone, 2013; Lee & Ronowick, 2014; Dewi, 2015), creativity (Polliastri & Paulina, 2009; Dewi, 2015), indecision (Doyle, 2003), problem solving (Hurtado,1995, as cited in Waddington, 2001; Schmitt, 1998; Nord, 2009; Dewi, 2015), revision (Ainsworth & Viegut, 2006; Washbourne, 2014; Dewi, 2015), time management/organization (Doyle, 2003; Dewi, 2015). Finally, based on the criteria introduced in the literature review and the instructors' opinions, a 28-item assessment checklist was developed. Then, to ensure the importance of the selected assessment criteria, the researchers asked the 12 instructors to answer the checklist consists of 28 items in "Yes/No" format (Appendix A). Only nine of the 12 instructors filled out the checklist completely. According to the instructors' point of view, 15 out of 28 criteria were merged into five[2], and five criteria including items (18) Initiative, (19) Pace of work, (21) Quality of terminological database, (22) CAT skills, and (23) Relevance of bibliography were added. Consequently, the analysis of the data showed that 23 translation assessment criteria were considered important to be used in the assessment rubric.

Therefore, the 23-item "Translation Quality Assessment Rubric" as the third instrument of this study was developed and validated to be applied for the assessment of students' performance on translation in the context of Iran (Appendix B).

The fourth instrument was a 607-word translation test included two texts (journalistic and political) taken from a news article and a textbook on political translation which were appropriate for translation courses at the BA level regarding the length and difficulty level. The BA translation students were invited to translate the texts from the English language into Persian.

---

[2] Items "5, 6, 10, and 11 in Appendix A were merged into Item 15 in Appendix B"; "Items 8 and 25 in Appendix A were merged into item 7 in Appendix B"; "Items 13, 23, and 26 in Appendix A were merged into item 16 in Appendix B"; "Items 16, 17, and 22 in Appendix A were merged into item 13 in Appendix B"; and "Items 19, 20, and 21 in Appendix A were merged into item 12 in Appendix B".

3.3. Procedure

The procedure undertaken in conducting the present research involved three phases as designing, implementing, and validating the assessment rubric. The first phase of the study was aimed at finding the expert translation instructors' perspectives over the criteria involved in assessing translated texts based on the data gathered from a 5-item researcher-made interview during October 2019. The interviews were held in a semi-structured form and performed in a conversational style in a single session. This kind of interview was flexible in posing extra required probing questions based on interviewees' earlier answers. In this type of interview, each answer was checked and the examinee was asked to give more explanations on their answers. They were given as much time as necessary to answer each question. They could express their thoughts verbally in their native language (Persian). All interviews were recorded and the interview time lasted almost 15 to 20 minutes in each session. The interviewers (the researchers) recorded all the interviews for later transcription and analysis. Then the researcher transcribed and analyzed the recorded interviews. Theme-based categorization (Dörnyei, 2007) was used to categorize the answers that emerged from the interview. The data were systematically structured and analyzed by classifying them through finding commonalities between them. In this regard, an in-depth analysis of interview data revealed that 32 assessment criteria were commonly introduced by 12 translation instructors.

During the second phase of the research, based on comparing the common assessment criteria that emerged from the review of the literature and those mentioned in interview answers, the translation assessment checklist was developed, including 28 items (Appendix A). The items were written in the English language. The assessment checklist includes the most common criteria involved in assessing translated texts. The 12 expert translation instructors were again invited to check the confirmation of the 28 criteria (Only nine of the 12 instructors participated in this part of the study). They answered the 28 items in a "Yes/No" question format. The instructors should determine whether the assessment criteria were appropriate for assessing a translated text by answering yes or no for each individual item. Consequently, the criteria which did not consider appropriate were removed, some were renamed, some were split or merged, and some changes were applied to the priority level of items. In the end, a 23-item translation assessment criteria were developed to use as the assessment rubric.

In the third phase of the research, 60 BA translation students, who were taking most of the technical translation courses on different subjects, were invited to translate two unseen texts from English into Persian (L2 into L1). Participants were asked to translate the texts at home and bring them to the class next week or returned it directly by email. They had the right to use any kind of resources such as a dictionary, internet, translator-aided tools and/or other software. No time constraints were imposed for translating the texts. The researchers regarded the translation test as a useful tool since it consisted of all factors included in the rubric to be assessed. The texts contained journalistic and political topics which studied during the translation courses at universities. The journalistic text was an extract of 303 words from a news article and the political text contained 304 words from a textbook on political translation. The political and journalistic text includes 11 and 12 statements, respectively. Therefore, both texts were of a similar length and difficulty and contained a similar range of linguistic issues.

The students were also asked to mention if they used any kinds of terminological databases, CAT (Computer –Aided Translation) tools, or the relevance bibliography for translating the text at the end of the translation test. Data collection in this phase was carried out from October 2019 to February 2020. Detailed instructions were provided to the translation students to make them evidently aware of the nature and intention of the test.

Finally, the developed 23-item translation assessment rubric was used to assess the quality of students' translation. The rubric was constructed on a four-point Likert scale ranging from "superior= score 4", "advanced= score 3", "fair= score 2", to "poor = score 1". Therefore, the minimum score for each item was 1, which means the lowest level of translation performance. In the same vein, the maximum scores for each item were 4, which means the highest level of translation performance. To ensure the inter-rater reliability of the scores, three raters, who were all experts in translation assessment, were assigned to evaluate the student's translation performance. Then, Rasch analysis was used to analyze the construct validity of the developed rubric for assessing the translation performance of the students.

4. Results

A test or questionnaire is said to be valid when the items' underlying construct causes the item responses (Baghaei & Tabatabaee-Yazdi, 2016). In view of that and to confirm the construct validity of the "Translation Quality Assessment Rubric", the data were subjected to Rasch analysis, which has been used widely to analyze the construct validity in social science research areas, using Winsteps software version 3.73 (Linacre, 2009).

4.1 Individual Item Characteristics
Table 6 shows the fit indices of the items. The "MEASURE" column shows the difficulty of the items, and "MODEL S. E." illustrates the standard error of the item difficulty measures. As it is shown by the Table, the items are set from difficult to easy. As Table 1 shows, Item 22 (CAT Skills) is the most difficult item on the rubric while the easiest item is Item 12 (Genre). Therefore, the difficulty of item 40 is estimated to be 1.72 logits with the standard error of 0.18, which means it can be 95% sure that the true value for the difficulty of this item positioned somewhere between 1.36 to 2.08 logits, i.e., two SE's below and above the observed measure.

Rasch analysis uses the separation index as the reliability indices (Linacre, 2009). Separation reliability shows "how well the person parameters are discriminated on the measure variable" (Tabatabaee-Yazdi, Motallebzadeh, Ashraf, & Baghaei, 2018, p. 134). Therefore, a high-reliability value signifies a strong association between the items of the test. Accordingly, the analyses of the items yield an item difficulty range of -1.65 to 1.72 logits with separation reliability of .96. Rasch person estimates ranged from -0.92 to 1.86, with separation reliability of .67. Thus, the study revealed to have an accepted medium reliability value. The study's four-point Likert scale, which might be considered as a short scale, could be the reason for this medium reliability index.

"INFIT" and "OUTFIT" mean square (MNSQ) should be ranged between 0.60–1.40 (Bond & Fox, 2007). Unusual response patterns that mislead the analysis and are defined as signs of construct irrelevant variance and multidimensionality (Baghaei, 2008) are shown by values

larger than 1.40. Values smaller than 0.60 do not mislead the analysis since they display redundancy of information. They can lead to false high reliabilities (Tabatabaee-Yazdi et al., 2018). According to these criteria, the results of Infit MNSQ columns, signify all the items are situated within the acceptable range of 0.60 to 1.40.

Table 6

Item Measures and Fit Statistics for the "Translation Quality Assessment Rubric"

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
22	166	80	1.72	.17	1.40	2.7	1.39	2.6
19	168	80	1.67	.17	1.22	1.5	1.20	1.4
18	173	80	1.53	.17	.81	-1.4	.81	-1.4
16	184	80	1.22	.17	.73	-2.0	.73	-2.0
17	185	80	1.19	.17	.94	-.4	.95	-.3
7	218	80	.25	.17	.89	-.7	.90	-.6
20	221	80	.16	.17	1.15	.9	1.13	.9
23	221	80	.16	.17	1.18	1.2	1.15	1.0
14	222	80	.13	.17	.78	-1.5	.76	-1.6
11	223	80	.10	.17	.69	-2.2	.69	-2.2
1	225	80	.04	.17	.64	-2.6	.64	-2.6
15	225	80	.04	.17	.64	-2.6	.64	-2.6
2	236	80	-.30	.18	1.13	.8	1.11	.7
13	236	80	-.30	.18	.74	-1.8	.73	-1.9
5	239	80	-.40	.18	1.08	.5	1.06	.4
21	239	80	-.40	.18	1.08	.5	1.06	.4
4	240	80	-.43	.18	1.21	1.3	1.19	1.2
3	247	80	-.65	.18	1.10	.7	1.09	.6
6	247	80	-.65	.18	.90	-.6	.87	-.9
10	257	80	-.99	.19	1.16	1.1	1.14	1.0
9	258	80	-1.02	.19	1.22	1.4	1.20	1.3
8	268	80	-1.38	.19	1.12	.9	1.12	.9
12	275	80	-1.65	.20	1.07	.5	1.07	.5
MEAN	224.9	80.0	.00	.18	1.00	-.1	.98	-.1
S.D.	30.3	.0	.92	.01	.22	1.5	.21	1.4

4.2 Response Scale Analyses

The category statistics for the 4-point scale was shown in Table 7. The category "observed average" was listed in the third column. This is the mean of all persons in the sample who chose that category. It is supposed that observed averages to increase with category values which is the pattern observed with these data.

The infit mean-squares and outfit mean-squares for each category level are the average of the infit and outfit mean-squares associated with the responses in each category. The values above 1.50 are considered problematic (Linacre, 2009). As it is shown in Table 2, all categories were within acceptable limits.

In evaluating rating scale models, "threshold" values should be in order. Disordered thresholds show that the category is not well-defined for respondents (Linacre, 1999), which means the respondents cannot clearly discriminate the scales (Bond & Fox, 2007). In this study, the thresholds column was shown to be in order (-2.15, -0.16, 2.31).
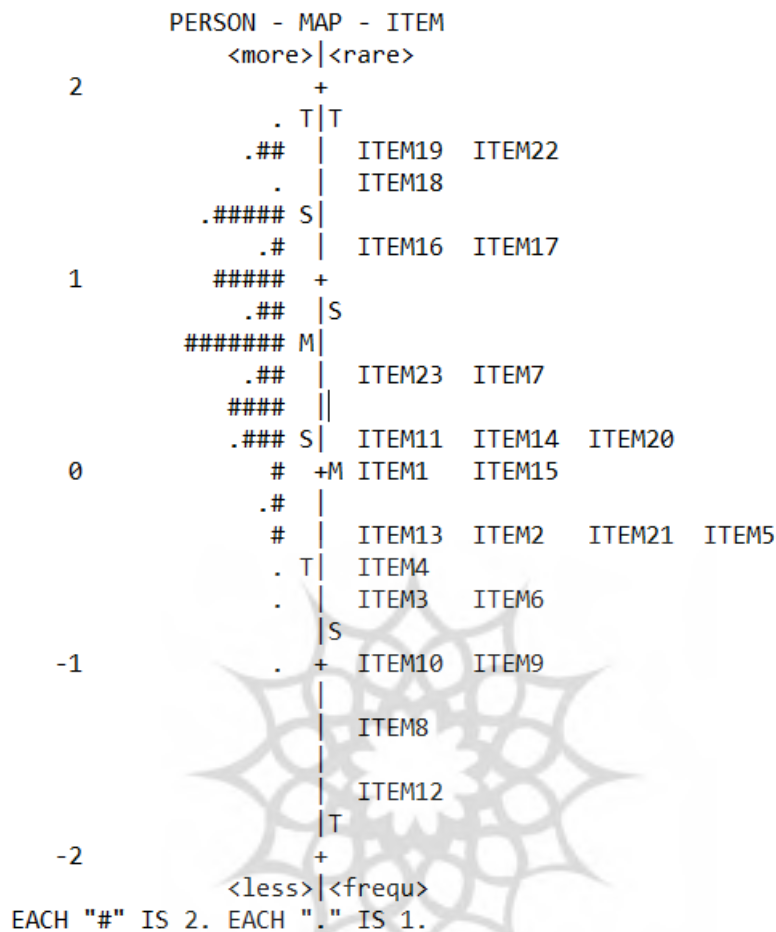
Table 7

Category Statistics

Category	Count (%)	Observed average	Infit MNSQ	Outfit MNSQ	Threshold	Category Measure
1 Poor	87 (5)	-.33	1.37	1.39	None	-3.34
2 Fair	496 (27)	-.27	0.71	0.70	-2.15	-1.18
3 Advanced	934 (51)	1.06	0.69	0.74	-0.16	1.11
4 Superior	323 (18)	1.40	1.20	1.17	2.31	3.47

Figure 1 illustrates a graphic depiction of the probability curves for the response categories. According to this curve, each category must have a peak on the curve to show a particular section of the measured construct. In this study, all the categories are shown to have a peak on the curve to specify a unique section of the measured construct.

Figure 1. Category Probabilities curve

Figure 2 shows the person-item map of the data. The person-item map shows the location of item parameters as well as the distribution of person parameters. According to Bond and Fox (2007), items should be perfectly spread along the scale to significantly measure all persons' abilities. Items located on top of the map are more difficult, and the ones down the scale are the easier and less proficient ones.

The person-item map revealed that the persons are mainly clustered toward the center and top of the scale, suggesting that persons with moderate and high translation skills are likely to answer most of the items correctly.

```
                PERSON - MAP - ITEM
                   <more>|<rare>
        2              +
                    . T|T
                    .## |   ITEM19  ITEM22
                     .  |   ITEM18
                .##### S|
                   .#  |    ITEM16  ITEM17
        1        #####  +
                   .## |S
              ####### M|
                   .## |    ITEM23  ITEM7
                 #### ||
                .### S|    ITEM11  ITEM14  ITEM20
        0          #  +M ITEM1   ITEM15
                  .# |
                   # |     ITEM13  ITEM2   ITEM21  ITEM5
                 . T|    ITEM4
                 .  |     ITEM3   ITEM6
                    |S
       -1          . +    ITEM10  ITEM9
                    |
                    |    ITEM8
                    |
                    |    ITEM12
                    |T
       -2           +
                 <less>|<frequ>
        EACH "#" IS 2. EACH "." IS 1.
```

Figure 2. Items-person map

4.3 Examination of Unidimensionality

To check the unidimensionality of the scale, principal components analysis (PCA) on standard residuals has been used and global fit statistics were studied by checking patterns in the residuals. The smaller the residuals, the better the data fit the model. It is supposed that the residuals are accidentally disseminated and uncorrelated (Linacre, 2009). If the latent trait enlightens all the info in the data and the residuals define random noise, the data is said to be fit by the Rasch model. Therefore, as Baghaei and Cassady (2014) stated, if a factor with size and strength of at least two items (eigenvalue=2) is extracted from the residuals, the test cannot be considered as unidimensional. Therefore, the size of the eigenvalue in the first factor is a degree of unidimensionality or overall fit of data to the Rasch model (Smith, 2002) which can be checked through the row "unexplained variance in the 1st contrast" in Table 3.

The eigenvalue of 1.9 for the first contrast in the present study shows that the test is unidimensional. PCA of the standardized residuals showed that the Rasch dimension is as big

as 18.3, which explains 85.4 % of the variance; 62.0 % is explained by item measures, and 23.4 % is explained by person measures. In all, 14.6 % of the variance remains unexplained.

Table 8
Dimensionality output

	Empirical		--	Modeled
Total raw variance in observations	24.1	100.0 %		100.0 %
Raw variance explained by measures	18.3	85.4 %		85.1 %
Raw variance explained by persons	5.1	23.4 %		23.1%
Raw Variance explained by items	13.2	62.0 %		62.0 %
Raw unexplained variance (total)	5.8	14.6%	100.0 %	14.9%
Unexplained variance in 1st contrast	1.9	6.4 %	9.3 %	

The study's overall findings of item characteristics and response scale quality confirm that the Rasch model fits the "Translation Quality Assessment Rubric", and the internal validity of the rubric was confirmed. The result showed that 23-item had an acceptable outfit mean-square and infit mean-square fit statistics. To sum up, the study's overall findings confirm that the 23-item of "Translation Quality Assessment Rubric" is an effective unidimensional representation for assessing the quality of translation performance.

5. Discussion

Due to the important role of research in designing translation assessment rubric for Iranian translation instructors and examiners and lack of a valid assessment rubric in the Iranian context an attempt was made to construct and validate a "Translation Quality Assessment Rubric" using the Rasch measurement model. The analyses were addressed the identification of evidence for a unidimensional structure for the rubric, and confirmation of the efficacy of a 4-point Likert type scale. The Rasch measurement model (Andrich, 1978) was used to validate the assessment rubric. Results of examination of response scales' item characteristics and the quality of responses confirmed that the Rasch model fits the "Translation Quality Assessment Rubric" well, which indicates the internal validity of the rubric. The rubric had an acceptable person separation reliability of .67 and item separation reliability of .96. Additionally, analyzing the hierarchy of item difficulties gives significant information about the subscales of the exam and specifically regarding current translation competencies. As items were set from difficult to easy, the person-item evaluation showed that item 22 *"CAT Skills"* was the most difficult item on the rubric to be endorsed, while the least likely to be endorsed item was item 12 *"Genre"*. Items that are easily endorsed and received higher scores emphasized the translation assessment criteria which were more effective in the translation quality assessment than other indicators. Whereas, those items which received lower scores demonstrated indicators that were less prominent to respondents or not fully observed by participants as appropriate translation assessment criteria.

To be suited for translation jobs, translation students are suggested to acquire some main translation competencies at universities which are important in doing translation tasks (Gómez González-Jover, 2011). Moreover, to have valid and reliable scores, assessment of learners' understanding of various concepts seems to be significant (Tabatabaee-Yazdi, 2020). Indeed, it seems to bridge the gap between translation practice, translation theory, and translation assessment (Lauscher, 2000; Snell-Hornby et al., 1994), the students at the BA level should acquire not only the necessary translation competencies but also translation assessment criteria which were introduced and approved by translation scholars and translation instructors in the field (Samir, Khoshsaligheh, 2012). The study results showed all items fit the Rasch model. This means that nearly all of the translation respondents at the BA level were familiar with various translation competencies associates with the assessment criteria in the rubric. Thus, to examine and monitor the professionalism of translation students and train qualified translators who use specific translation competencies and techniques, such objective criteria should be assessed by the rubric.

6. Conclusion

Choosing an appropriate translation assessment methods and criteria are ideal tasks for translation instructors and raters in Translation Studies. Many translation quality assessment methods have been proposed and some of them were implemented in practice. However, the crucial concern in successful assessment is the availability of a valid and reliable rubric that corresponds with learning needs and instructors' perspectives. In addition, evaluating the quality of translation by using a valid and reliable rubric is essential for a successful education. However, few reliable and valid assessment rubrics have been developed for assessing the quality of translation specifically in the Iranian higher education system. In light of these explanations, this study aimed to construct and validate a "Translation Quality Assessment Rubric" using the Rasch measurement model.

To conclude, the study's overall findings show that the 23-item Translation Quality Assessment Rubric is an effective unidimensional representation for assessing the quality of students' translation performance. The findings reveal that these assessment criteria are effective in practice for objective assessment of students' translation performance. Thus, the findings could provide major practical implications for translation instructors, raters, and students. The development of the Translation Quality Assessment Rubric could give much support to translation instructors, who are the main implementer of this assessment rubric, by helping them to become familiar with how to improve the translation quality and how to evaluate the translated texts. Besides, the familiarity of the translation students with these assessment criteria could assist them in evaluating and assessing their own translations in any text type.

References

Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Corwin Pr.

Akbari, A., & Shahnazari, M. (2015). The rejuvenation of equivalence paradigm in literary texts: Equimediation as the seul et alleiniges paradigm in Translation. *Mediterranean Journal of Social Sciences*, *6*(4), 438-438.

Akhiroh, N. S. (2013). The influence of translation technique on the quality of the translation of international news in Seputar Indonesia daily. *Language Circle: Journal of Language and Literature*, *7*(2).

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Angelelli, C. V. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V.Angelone, E. (2013). The impact of process protocol self-analysis on errors in the translation product. *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association*, *8*(2), 253-271.

Arango-Keeth, F., & Koby, G. S. (2003). Assessing assessment: Translator training evaluation and the needs of industry quality assessment. In *Beyond the Ivory Tower* (pp. 117-134). John Benjamins.

Bachman, L. F. (2014). Ongoing challenges in language assessment. In A. J. Kunnan (Ed.), The companion to language assessment, Volume 3, 1586-1603. Oxford: John Wiley and Sons.

Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests, Volume 1. New York: Oxford University Press.

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement Transactions*, *22*, 1145-1146.

Baghaei, P., & Cassady, J. (2014). Validation of the Persian translation of the Cognitive Test Anxiety Scale. *Sage Open, 4*, 1-11.

Baghaei, P., & Tabatabaee Yazdi, M. (2016). The logic of latent variable analysis as validity evidence in psychological measurement. *The Open Psychology Journal. 9*, 168175.

Baker, M., & Saldanha,G .(Eds.)(2009). Routledge Encyclopedia of Translation Studies. Newyork: Routledge.

Beeby, A. (2000). Teaching translation from Spanish to English. Ottawa: University of Ottawa Press.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Broeck, R. van den. (1978). The concept of equivalence in translation theory. Some critical reflections. In Holmes, J.S., Lambert, J. and Broeck, R. van den (eds.) *Literature and Translation*. Leuven: Academic, 49-58.

Campbell, S., & Hale, S. (2003). Translation and interpreting assessment in the context of educational measurement. Translation Today: Trends and Perspectives 14, 205-224.

Cohen, A. D. (1994). Assessing language ability in the classroom. Boston: Heinle & Heinle Publishers.

Conde, T. (2013). Translation versus language errors in translation evaluation. In D. Tsagari & R. vam Deemter (Eds.) *Assessment Issues in Language Translation and Interpretation* (pp. 97-112). Peter Lang.

Conde, T. (2012). Quality and quantity in translation evaluation: A starting point. *Across Languages and Cultures*, *13*(1), 67-80.

Dewi, H. D. (2015). *Comparing two translation assessment models: Correlating student revisions and perspectives* (Doctoral dissertation), Kent State University.

Doyle, M. S. (2003). Translation pedagogy and assessment: Adopting ATA's framework for standard error marking. *The ATA chronicle*, *32*(11), 21-28.

Dweik, B. S., & Suleiman, M. (2013). Problems encountered in translating cultural expressions from Arabic into English. *International Journal of English Linguistics*, *3*(5), 47.

Farahzad, F. (1992). Testing achievement in translation classes. In C. Dollerup & A. Loddegaard (Eds.), *Teaching translation and interpreting: training, talent, and experience* (pp. 271-278). Amsterdam/Philadelphia: John Benjamins.

Genesee, F. (2002). Evaluation. In R. Carter, & D. Nunan (Eds.), The Cambridge Guide to Teaching English to Speakers of Other languages (4th ed., pp. 144-150). Cambridge: Cambridge University Press.

Genesee, F., & Upshur, J. (1996). Classroom-based Evaluation in Second Language Education. Cambridge: Cambridge University Press.

Gharacheh, B. (2005). The effect of back translation on the assessment of translation. (Unpublished Master Thesis), Tehran Azad University.

Goff-Kfouri, C. A. (2005). Testing and evaluation in the translation classroom. *Translation Journal, 9*(2), 75-99.

Gómez González-Jover, A. (2011). Course design and lesson planning in legal translation training. *Perspectives, 19*(3), 253-273.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Baltimore, MD: University Park Press.

Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in English*. Routledge.

Hanifehzadeh, S., & Farahzad, F. (2016). Development and validation of a researcher constructed psycho-motor mechanism scale for evaluating the quality of translation works. *International Journal of Language Testing, 6*(2), 72-91.

Hatim, B. & Mason, I. (1997). The Translator as Communicator. London, U.K. & New York, N. Y.: Routledge.

Hansen, G. (2010). Translation 'errors'. *Handbook of translation studies*, *1*, 385-388.

Heaton, J.B. (1990). *Classroom testing*. New York: Longman.

House, J. (1997). Translation quality assessment: A model revisited. Nehren: Gunter NarrVerlag Tubingen.

House, J. (2015). Translation quality assessment: Past and present. London and New York: Routledge.

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational research*, *60*(2), 237-263.

Khorami, F., & Modarresi, G. (2019). A Rasch-based Validation of the Evaluation Rubric for Consecutive Interpreting Performance. *Sendebar: Revista de la Facultad de Traducción e Interpretación*, (30), 221-244.

Kim, M. (2009). Meaning-oriented translation assessment. In C. V. Angelelli & H. E.Jacobson, Testing and assessment in translation and interpreting studies, 123157. Amsterdam: John Benjamins.

Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. Manchester: St Jerome Publishing.

Kelly, D. (2008). Training the trainers: towards a description of translator trainer competence and training needs analysis. *TTR: traduction, terminologie, rédaction, 21*(1), 99-125.

Khanmohammad H., O., M., & Osanloo, M. (2009). Moving toward objective scoring: A rubric for translation assessment. *JELS, 1*, 131-153.

Koby, G. S., Fields, P., Hague, D. R., Lommel, A., & Melby, A. (2014). Defining translation quality. *Tradumàtica*, (12), 0413-420.

Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet?. *The translator*, *6*(2), 149-168.

Lee, Y. O., & Ronowicz, E. (2014). The development of an error typology to assess translation from English into Korean in class. *Babel*, *60*(1), 35-51.

Lei, M. U. (2006). Translation competence and translation testing. *Shanghai Journal of Translators*, 8(*2*), 59-63.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 103-122.

Linacre, J. M. (2009). *A user's guide to WINSTEPS*. Chicago, IL: Winsteps.

Mackenzie, R. (2004). The competencies required by the translator's roles as a professional. In Malmkjær, K. (Ed.), *Translation in undergraduate degree programs* (pp. 32-38). Philadelphia, PA: John Benjamins Publishing Company.

Mahmoudian, S., Shahmiri, E., Rouzbahani, M., Jafari, Z., Keyhani, M. R., Rahimi, F., & Farhadi, M. (2011). Persian language version of the. *The International Tinnitus Journal*, *16*(2), 93-103.

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. Practical Assessment, Research & Evaluation 7 (25), 1-10.

Modarresi, G., & Ghoreyshi, S. V. (2018). Student-centred corrections of translations and translation accuracy: A case of BA translation students. *Translation Studies Quarterly*, *15*(60).

Moss, A., & Holder, C. (1988). Improving student writing. Dubuque, IO: Kendall/ Hunt.

Munday, J. (2001). Introducing Translation Studies: Theories and Applications. London: Rotledge.

Muñoz, M. R. (2012). *Cognitive and psycholinguistic approaches.* In M. C. Millán & F. Bartrina, Handbook of translation studies (pp. 241–256). London: Routledge.

Muzii, L. (2007). Adaptation, translation, and multimediality. *Online Translation Journal, 1*(4), 46-58.

Neubert, A. (2000). Competence in language, in languages, and in translation. *Benjamins Translation Library*, *38*, 3-18.

Newmark, P. (1988). A textbook of translation. New York: Prentice-Hall International.

Nord, C. (1999). Translating as a text-production activity. *On-Line Symposium on Innovation in Translator and Interpreter Training, Interpreter Training*, 3(4), 11-19.

Nord, C. (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis* (No. 94). Rodopi.

Nord, C. (2009). *Text analyse und Űbersetzen* (*text analysis and translation*). Tübingen: Gunter Narr.

Ørsted, J. (2001). Quality and efficiency: Incompatible elements in Translation Practice?. *Meta: journal des traducteurs/Meta: Translators' Journal*, *46*(2), 438-447.

Polliastri, P., & Paulina, A. (2009). Evaluation criteria for the improvement of translation quality. *Enhancing translation quality: ways, means, methods*, 239-261.

Qomariah, L. R., & Thahara, Y. (2015). The analysis of personification translation in the novel of Angels and Demons. *Unpublished Thesis. Situbondo: Abdurachman Saleh Situbondo University*.

Rea-Dickins, P., & Germaine, K. (1993). Evaluation. Oxford: Oxford University Press.

Reiss, K. (2014). *Translation criticism-potentials and limitations: Categories and criteria for translation quality assessment*. Routledge.

Riazi, A. M. (2003). The invisible in translation: The role of text structure. The Translation Journal, 7(2), 1-8.

Rico-Perez, C. (2002). Translation and project management. *Translation Journal, 6*(4), 38-52.

Schaffner, C. H. R. I. S. T. I. N. A., & Adab, B. (2000). Developing translation competence: Introduction. *Benjamins Translation Library*, *38*, vii-xvi.

Sager, J. C. (1983). *Quality and standards: The evaluation of translations*. In C. Picken (Ed), The translator`s handbook (pp. 91102). London: ASLIB.

Samir, A., & Khoshsaligheh, M. (2012). Translation students 'opinions, comments, and feedback on TQA approaches of translator trainers. *Journal of language and Translation Studies, 45* (3), 1-21.

Sharififar, M., Beh-Afarin, S. R., & Younesi, H. (2018). Classroom translation assessment techniques: How can we tell what/how our students are translating?. *International Journal of Language Testing, 8*(2), 44-53.

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*, 205-231.

Snell-Hornby, M., Pochhacker, F., & Kaindl, K., (Eds). (1994). *Translation Studies: An Interdiscipline* (Benjamins Translation Library 2). Amsterdam & Philadelphia: John Benjamins Publishing.

Stansfield, C. W., Scott, M. L., & Kenyon, D. M. (1992). The measurement of translation ability. *The Modern Language Journal*, *76*(4), 455-467.

Stobart, G., & Gipps, C. (1997). Assessment: A teacher's guide to the issues (3rd ed.). London: Hodder & Stoughton

Stevens, D. D., & Levi, A. (2004). *Introduction to Rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning. Stylus Pub LIC*. Retrieved from http://catalog.ebay.ca/introductionto-rubricsins.

Tabatabaee-Yazdi, M. (2020). Hierarchical Diagnostic Classification Modeling of Reading Comprehension. *SAGE Open*, *10*(2), 2158244020931068.

Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and Validation of a Teacher Success Questionnaire Using the Rasch Model. *International Journal of Instruction*, *11*(2), 129-144.

Techno Translation Studio. (2016, January 14). Translation quality: A customer's guide. Retrieved from https://technolex.com/articles/translation-quality.html

Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (Vol. 11). John Benjamins Publishing.

Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta: journal des traducteurs/Meta: Translators' Journal*, *46*(2), 311-325.

Walvood, B. E., & Anderson, V. J. (1998). Effective grading. San Francisco: Jossey-Bass.

Washbourne, K. (2014). Beyond error marking: written corrective feedback for a dialogic pedagogy in translator training. *The Interpreter and Translator Trainer*, *8*(2), 240-256.

Wiggins, G. (1998). Educative assessment: Designing assessments to inform and improve student performance. London: Jossey-Bass Publishers.

Williams, M. (2001). The application of argumentation theory to translation quality assessment. *Meta: Translators' Journal*, *46*(2), 326-344.

Williams, M. (2009). Translation quality assessment. Mutatis Mutandi, 2 (1), 3-23

Wilss, W. (1998). Decision making in translation. In M. Baker (Ed.), Routledge encyclopedia of translation studies (pp. 57-60). London/New York: Rutledge.

Appendix A

Translation Quality Assessment Checklist
Dear respondent
This checklist is designed to investigate the translation assessment criteria the instructors consider important in assessing the quality of translation. Your responses will be used by researchers to develop a translation quality assessment rubric for assessing student's translation. Your careful completion of the checklist will definitely contribute to obtaining real data, which is crucial for more accurate findings. We are really thankful for your real responses. The information you reveal to us through this checklist is kept confidential and would be used only for research purposes.

Assessment Criteria	Determined by the Translation Instructors	
	Yes	No
1. Do you assess the quality of target language **grammar**?		
2. Do you assess the idiomatic **usage** of the target language?		
3. Do you assess the quality of **omission** of something from the TT which is unessential to the meaning?		
4. Do you assess the **accuracy** of transferring the content and meaning from the ST into TL?		
5. Do you assess the quality of the **naturalness** of language form in translation?		
6. Do you assess the **ambiguous** words and expressions that leads to vagueness in TT?		
7. Do you assess the quality of the **addition** of something in the TT which is not clearly expressed in the ST?		
8. Do you assess the quality of the translation of specific **terms** embedded in technical text?		
9. Do you assess the usage of the TL **punctuation** convention in the TT?		
10. Do you assess the quality of TT **fluency**?		
11. Do you assess whether the **mistranslation** of meaning in TT occur?		
12. Do you assess the quality of **faithfulness** of translation?		
13. Do you assess whether the translator's **indecision** about meaning or grammar in the TT?		
14. Do you assess the appropriate use/preserve of **register** in the translation?		
15. Do you assess whether all sections within the ST transfer **completely** to TT?		
16. Do you assess the quality of **cohesion** in the TT?		
17. Do you assess the quality of **coherence** in the TT?		
18. Do you assess the **italicization** and **capitalization** of words and expressions in TT?		
19. Do you assess the appropriate use/preserve of source language **text style** in the TT?		
20. Do you assess the quality of retention of the **genre** in the TT?		
21. Do you assess the appropriate use/preserve of source language **text type** in the TT?		
22. Do you assess the quality of lexical **consistency** in translation?		
23. Do you assess the quality of **creativity** used in the TT?		
24. Do you assess the **spelling** of each word in the TT?		
25. Do you assess the quality of **false friends'** equivalent in TT?		
26. Do you assess the translator's **problem-solving** skills?		
27. Do you assess the quality of the **revision**/**self-assessment** technique used in the TT?		
28. Do you assess the **time management skills** of the translator?		
Total Score		

Appendix B

Translation Quality Assessment Rubric
Dear respondent
This Rubric is developed to assess the overall quality of student's translation in Iranian universities. Your rate will be used by researchers to determine the quality of the student's translation. To that end, your careful completion of the rubric will definitely contribute to obtaining real data, which is crucial for more accurate findings. The information you reveal to us is kept confidential and would be used only for research purposes.

A. Please mark the part that best matches your status.
Gender: Male ☐ Female ☐ City:………………
Your level of qualification: PhD ☐ M.A. ☐
Years of translation experiences: Below 5 years ☐ 5-10 years ☐ more than 10 years ☐
Field of teaching: TEFL ☐ English Translation ☐ English Literature ☐ Others:.............

B. Respond to each statement below according to the following 4-point Likert scale:
Superior = Score 4, Advanced= Score 3, Fair = Score 2, and Poor = Score 1

Students Code: ………………

Items	Rate by the Raters			
	1	2	3	4
1. Grammar (Word Form/ Part of Speech, Word Order, Syntax …)				
2. Usage				
3. (No) Addition				
4. (No) Omission				
5. Completeness				
6. Punctuation				
7. Terminology/False friend Terminology				
8. Spelling				
9. Capitalization/ Italicization Rules				
10. Faithfulness/literalness				
11. Register/Tone				
12. Genre (Text Style, Text Type)				
13. Cohesion/Coherence, Consistency				
14. Accuracy				
15. Fluency (Naturalness, Readability, No Ambiguity, No Mistranslation)				
16. Creativity/Problem Solving (No Indecision)				
17. Organization/time management				
18. Initiative				
19. Pace of work				
20. Revision file, self-assessment				
21. Quality of terminological database				
22. CAT skills				
23. Relevance of bibliography				
Total				

The Definitions of the key translation assessment criteria:

1. Grammar: There has to be an agreement between subject and verb. Therefore, the translator should focus on the agreement and correct use of verb forms, adjectives, adverbs, nouns, and pronouns. Translators also must use the correct form of the word. For instance, a word must be a noun but it is written in verb (i.e. *to weaken* instead of *to weak*1). Word order is different in different languages and in translation, a modification is required. The order of SOV and modifiers in the English language can be changed in TL. For example, the word order in *she loves him* (SOV) in the English language is changed into *loves she him* (VSO) in Welish language. Therefore, words arrangement in a sentence or clause should follow the rules of the TL, and the translator must use correct syntax so as not to change the meaning.

-Does the translator follow grammatical rules such as subject and verb agreement? Does the translator appropriately use the relative order of subject, verb, modifiers, clauses, and syntactic elements? Does the translator appropriately arrange words according to the TL rules? Does the translator use the correct form of the word?

2. Usage: Appropriate use of target language. In this case, the translator should focus on the correct use of prepositions and/or grammatical form for instance worry about versus worry for, do shopping versus perform shopping.

-Does the translator produce correct and idiomatic usage of the target language?

3. Addition Add something that did not express clearly in the ST but it was essential to the meaning.

-Does the translator insert something in the TT which is not clearly expressed in the ST?

4. Omission: Omit something that was not essential to the meaning.

-Does the translator leave out something from the TT which is unessential to the meaning?

5. Completeness: All sentences, paragraphs, titles, headings, subheadings must be translated.

-Does the translator convey all sections (sentences, titles, headlines …) within the ST to TT?

6. Punctuation: Appropriate use of TL punctuation such as those governing full stops, question marks, commas, colons, semi-colons, exclamation marks, quotation marks, dash, hyphen, parentheses, brackets, braces, slash, and paragraphing.

-Does the translator follow the conventions of the TL governing the use of quotation marks, semicolons, etc.?

7. Terminology: Correct use of technical terms with specific meanings used in legal, political, business, and economic texts, etc. In addition, the translator should be aware of the pairs of words names false friends in the two languages which drive from similar roots and have similar forms, but that have different definitions. For instance, Persian *kar* and English *car*.

-Does the translator use specific terms in the translation of the technical text? Does the translator translate appropriately the words which are derived from similar roots and have similar forms but have different meaning?

8. Spelling: There must be correct spelling to avoid confusion and misunderstanding with the aimed meaning in the context for instance bite and beet, especial and special, guess and guest, quiet and quiet.

-Does the translator avoid spelling error which causes misunderstanding about the intended meaning?

9. Capitalization/ Italicization: The main TL rules regarding italicization and capitalization must be strictly followed. For instance, in the Persian language, the first letters of the proper nouns are not capitalized.

-Does the translator follow the conventions of the TL in (not) carrying English capitalization and italicization of proper nouns into Persian?

10. Faithfulness: Not only the translator must focus on the denotative meaning and connotative meaning, but also they should focus on the cultural associations of meaning. It means transferring the intention of the author and the function of the ST appropriately.

-Does the translator very closely translate the message and the structure of the ST expression to the TT?

11. Register: Preserving appropriately the level of formality and/or language level in the translation. For instance, do not make the medical text sound journalistic or use legal terms instead of medical terms.

-Does the translator use/preserve an appropriate register (language level, degree of formality) in the translation?

12. Genre: The specific type of writing in a literary genre (poetry, fiction, nonfiction, and drama), journalistic genres, economic genre, and political genre, etc., with each differing in subject matter, structure, style, and the use of figurative language must reflect in translation. The particular manner of expression in ST for example conversational and literary words and expression must be retained in TT appropriately (i.e. style).

-Does the translator reflect the ST distinctive manner of expression in the TT? Does the translator produce TT according to the ST text type?

13. Cohesion/Coherence, Consistency: the words which were written in the ST consistently must be rendered in the TT consistently (same terms in ST must translate similarly in the TT)**.** Although coherence is pragmatic and rhetorical, cohesion is semantic and grammatical. Coherence and cohesion (e.g. linkages, appropriate use of pronouns, transitional devices, etc.) are closely tied. The coherent text is written by correct particles, connectors, reference to phenomena, temporal cohesion, and category. The translator must use the markers which have a role in making pragmatic coherence and semantic cohesion. Therefore, to develop a coherent text and to connect the textual element in paragraphs or between paragraphs effectively, the translator must use cohesive devices accurately.

-Does the translator use cohesive devices appropriately? Does the translator translate consistently the term that is used consistently in the ST?

14. Accuracy: the translator must transfer the content and meaning of the ST into TL completely.

-Does the translator demonstrate content and meaning at a good level of accuracy? Does the translator understand the words or syntax of the ST sentence?

15. Fluency (Naturalness, Readability, No Ambiguity, and No Mistranslation): The translator should use the natural form of TL. The translation should be written in ordinary TL grammar, words, idioms, word order, and syntax. In addition, the meaning in TT should not be ambiguous. In a misunderstanding, there is an error in TT because the translator did not understand or interpret the ST or misread the words or the syntax of the sentence. In

mistranslation, there is an error in TT because the translator did not convey accurately and completely the meaning of the ST in the TL.

-Does the translator produce the TT at an acceptable level of fluency? Does the translator avoid words and expressions having ambiguous meanings? Does the translator convey properly the meaning of the ST in the TT?

16. Creativity/Problem Solving (No Indecision): The translator should discover suitable equivalents in TL creatively. S/he must translate according to the text type, and follow the purpose of the translation. The TT should not have a minor problem so the translator must try to find an adequate solution to translation problems (i.e. find a creatively correct and clear word and expression in TT). S/he must not write more than one option for word and expression in TT, as well.

-Does the translator creatively discover proper equivalents? Does the translator leave the reader to choose between possible options?

17. Organization/time management: Translators should manage the time during the translation appropriately.

-Does the translator manage the time appropriately?

18. Initiative: there are some times when the translator faces some problems in translating so s/he takes the initiative to ask for help.

-Does the translator take the initiative to ask for help when required?

19. The pace of work: The translator should respect the translation deadline.

-Does the translator respect the deadline?

21. Quality of terminological database: The translator can use an appropriate terminological database during translation.

-Does the translator follow the instructions for the revision of the text?

-Does the translator use an appropriate terminological database?

22. CAT skills: Specialized translators have CAT (Computer-Aided Translation) skills to translate various technical texts. The translator should use translation memory, Termbases, machine translation engine, Autosuggest dictionary, or SDL Trados (Software and Documentation Localization), etc., to support the translation process and edit the TT.

-Does the translator have CAT skills?

23. The Relevance of bibliography: To discover appropriate definitions in TL, the translator can use other relevant bibliographies.

-Does the translator use a relevance bibliography?