

Moving toward Objective Scoring: A Rubric for Translation Assessment

Hajar Khanmohammad
Maryam Osanloo

Abstract

Setting standards for translation assessment has usually been problematic for authorities, theoreticians, and most commonly instructors. It is often difficult to tell the difference between translation evaluation, criticism, and assessment due to hardships of bridging the gap between theory and practice. The aim of this study was to create a rubric against which translations could be assessed. To this end, it sought to elaborate assessment aspects of translation with the hope to reduce rater subjectivity in scoring, and define exactly what factors should be taken into account in assessing translated texts by developing a new model for the evaluation and scoring of translations. Accordingly, the existing rubrics for the assessment and evaluation of translations were reviewed and subsequently, two questionnaires – one open-ended and one using Likert scale – were administered among a total of 41 translation instructors inquiring about their methods for assessing students' translations. Based on the results obtained from both questionnaires and also the existing scales of assessment, a rubric was developed to delineate a framework of translation assessment criteria.

Keywords: translation assessment, rubric, assessment criteria, translation accuracy, translation appropriateness

Introduction

The world today seems to be getting smaller and smaller as communication and information systems are developing and becoming more and more sophisticated. In the process of such a rapid exchange of information and for the purpose of improving cultural contacts, translating becomes inevitable. Once, translation was undertaken by individuals for individuals. However, increasing global demands prompted businesses seek markets worldwide, which generated further needs for translation of all types. While some scholars like Catford (1965), Savory (1968), Nida (1984), and Bell (1991) put emphasis on equivalence in translation, for others like Al-Qinai (2000),

translation is a complex hermeneutic process in which intuition plays a crucial role in interpreting the intentions of the source text writer, thus they argue that translation is a quality and quality is relative, and absoluteness of accuracy ceases where the end user imposes his/her own subjective preferences of style in translation.

Regardless of the theoretical camp one belongs to, no one would dispute over the importance of needing standards for evaluation and assessment of translation. Some scholars such as Bassnet (1980), Hatim and Mason (1997), Belloc (1931), and Gerasimov (2005) have tried to develop models that satisfy the needs of practitioners to bridge the gap between theory and practice. Others including Goff-Kfoury (2005) and Steven and Levi (2004) have attempted to draw up objective translation assessment criteria such as validity, reliability, and objectivity into the overall structure of the translation. In a practical sense, however, one of the simplest yet most fundamental requirements is that a translation should be assessed. This is of course especially true in the educational setting as well.

The first step in translation assessment is to establish a model of quality and then to transform it into a set of metrics that measure each of the elements of that quality. Muzii (2007) believes that a comprehensive set of criteria must assess the quality of translation from several perspectives during the production process. Making a single, all encompassing metric is not only troublesome, it is ironically likely to be useless since a simple metric would not reveal all the problems. Creating multiple metrics that assess the various aspects of what is to be measured – in this case translation – can help recompose the overall framework and give an indication of which parts of a process work well and which part does not. Hence, a reliable and valid rubric in translation assessment would be required to address the aforementioned issues.

Rubric, which etymologically refers to decorative text or instructions in medieval documents penned in red ink and later on in academia, referred to notes that a teacher penned in red ink while grading a paper has come to refer to a scoring tool in modern academic jargon. Rubrics are often supported and used in alternative assessments in education (Virginia Education Association and Appalachia Educational Laboratory, 1992; McTighe, 1997; Wangsatorntankhun, 2007) but have gained ground as a way of establishing written guidelines or standards of assessments for formal, professionally-administered essay tests. Observing how rubrics are used in academic settings and adopting and adapting rubrics for use in the many functional areas of student affairs, Stevens and Levi (2004) write, "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” (p. 3).

Student affair professionals constructing the rubric would describe within each dimension “what constitutes each level of performance” (Stevens & Levi 2004, p. 6). In essence, a rubric provides a means to score student performance vis-à-vis the learning and development objective and provides rich feedback on the level and characteristics of students’ present learning relative to that objective.

Riazi (2003) describes the rubric as an attempt to delineate consistent assessment criteria. He emphasizes that it allows teachers and students alike to assess criteria which are complex and subjective and also provide ground for self-evaluation, reflection, and peer review. It is aimed at accurate and fair assessment, fostering understanding, and indicating the way to proceed with subsequent learning/teaching. This integration of performance and feedback is called ongoing assessment. Increasingly, instructors who rely on rubrics to evaluate student performance tend to share the rubric with students at the time the assignment and exams are made. In addition to helping students understand how the assignment relates to course content, a shared-rubric can increase student authority in classroom through transparency.

Existing Rubrics for Translation Assessment

There are of course various rubrics in the literature of translation studies. The earlier ones were based on error analysis, examples of which include the one designed by Sager (1983) which was based on three types of errors of linguistic, semantic, and pragmatic elements in five levels for scoring, and the one designed by Mason (1997) in four levels of errors for every single sentence. Rubrics developed later, however, focused on positive points in addition to penalties, similar to the one developed by Rico Perez (2002) which was designed to detect six types of errors and three degrees of seriousness.

In the following section, only those rubrics considered as being practical in the literature are described, albeit briefly. These rubrics include the ones by Farahzad (1992), Waddington (2001), Sainz (1992), Beeby (2000), and Goff-kfoury (2005).

Farahzad's Rubric

Farahzad (1992, p. 274) maintains that two main features are to be checked in scoring for each unit of translation:

1. *Accuracy*: the translation should convey the information in the source text (ST) precisely, i.e. the translation should be close to the ST norms.
2. *Appropriateness*: the sentences should sound fluent and native, and should be correct in terms of structure.

She declares that unnatural translations which convey the ST's meaning receive half a score, whereas inaccurate translations receive no score, no matter how appropriate and natural the target texts sound. In error recognition items, one score is given for spotting the error and another one for correcting it. Farahzad (1992) believes that scoring a long text can be done in two different ways:

A: It can be scored holistically. Since the item assesses a wide variety of competencies, the examiner may find it convenient to approach the text as the unit of translation and adopt this system, especially with a large number of students. The examiner may, for instance, come up with the scheme as demonstrated in Table 1.

Table 1 – A rubric based on the text as the unit of translation (Farahzad, 1992)

Items	Score
Accuracy	<i>20 percent</i>
Appropriateness	<i>20 percent</i>
Naturalness	<i>20 percent</i>
Cohesion	<i>20 percent</i>
Style of	<i>20 percent</i>
<i>discourse/choice of words</i>	

B: It can be subjected to objectify scoring. In this system, the target text must be read two times, first to check the accuracy and appropriateness, then for cohesion and style (the details of which appear in Table 2 below). Although time-consuming, this system is more reliable. Farahzad continues

that sentence and clause might be the units of translation. Thus each verb in the source language text marks a score. The main clause receives one score and each sub-clause another score.

Table 2 – Farahzad's model based on sentence and clause as the unit of translation

Accuracy and appropriateness			Cohesion and style				
sentences	in clause	in clause	linkages	appropriate use of pronouns	choice of words	grammatical structures	stylistic devices

Cohesion and style cannot be checked and scored at the sentence and clause level. The elements of cohesion (e.g. transitional, appropriate use of pronouns, linkages, etc.) are spread all over the text as are the elements which form the style of discourse (e.g. choice of words, grammatical structures, etc.). If, for instance, the source text is fairly neutral, one may allot a smaller number of points to it than in other cases where the preservation of style is important.

Waddington's Rubric

In Waddington's (2001) model of translation quality assessment, each translated text is assessed by the researcher and two other raters. He declares that almost all the contributions in translation quality assessment

have been descriptive or theoretical and have centered mainly on the following themes:

- (i) Establishing the criteria for a 'good translation'
- (ii) The nature of translation errors:
 - Defining the nature of translation errors as opposed to language errors;
 - Drawing up a catalogue of possible translation errors;
 - Establishing the relative, as opposed to absolute, nature of translation errors; and
 - The need to assess quality not only at the linguistic but also the pragmatic level.
- (iii) Basing quality assessment on text linguistic analysis
- (iv) Establishing various textual levels on a hierarchical basis and linking the importance of mistakes to these levels
- (v) Assessment based on the psycholinguistic theory of "scenes and frames"

Waddington introduces four methods of assessment. The first method (Method A) is more well-known than the other three and is functional in translation classes. This method is based on error analysis and possible mistakes are grouped under the following headings:

- (i) Inappropriate renderings which affect the understanding of the source text and are divided into eight categories: counter-sense, faux sens, nonsense, addition, omission, unresolved extra-linguistic references, loss of meaning, and inappropriate linguistic variation (register, style, dialect, etc.).
- (ii) Inappropriate renderings which affect expression in the target language and are divided into five categories: spelling, grammar, lexical items, text, and style.
- (iii) Inadequate renderings which affect the transmission of either the main function or secondary functions of the source text.

In each of the categories, a distinction is made between serious errors (-2 points) and minor errors (-1 point). There is a fourth category which describes the plus points to be awarded for good solutions (+1 point) or exceptionally good solutions (+2 points) to translation problems. Table 3 displays such a scoring procedure.

Waddington (2001, p. 314) describes Method B as being based on error analysis and designed to take into account the negative effect of errors on the overall quality of the translations where the rater first has to determine whether each mistake is a translation mistake or just a language mistake. This is done by deciding whether or not the mistake affects the transfer of meaning from the source to the target text, and if it does not, it is a language error (and is penalized with -2 points). However, in the case of translation errors, the rater has to judge the importance of the negative effect each error has on the translation, taking into consideration the objective and the target reader specified in the instructions to the candidates in the exam paper. In order to judge this importance, Table 4 is suggested to the rater.

Table 3 – Serious and minor errors in Waddington's Method A

	Omission
Inappropriate rendering on understanding ST	Addition
	Nonsense
	Faux sens
	Counter-sense
	Style
	Text
Inappropriate rendering on TL	Lexicon
	Grammar
	Spelling
	Main function of
Inadequate rendering	ST
	Secondary
	function of ST
Good solutions	+1 point
	+2 points

Table 4 – Typology of errors in Waddington’s Method B
Negative effect **Penalty for**

on the words in ST	negative effect
1-5 words	1-5 points
6-20 words	6-20 points
21-40 words	21-40 points
41-60 words	41-60 points
61-80 words	61-80 points
81-100 words	81-100 points
100+ words	100+ points
<i>The whole text</i>	<i>The whole points</i>

In describing Method C, Waddington (2001, pp. 314-5) believes that this third procedure is a holistic method of assessment. The scale is unitary and treats the translation competence as a whole, but requires the rater to consider three different aspects of the student’s performance, as shown in Table 5 below. For each of the five levels, there are two possible marks; this allows the rater freedom to award the higher mark to the candidate who fully meets the requirements of a particular level and the lower mark to the candidate who falls between two levels but is closer to the upper one.

Table 5 – Scale for the holistic Method C (Waddington, 2001)

Accuracy	Quality	D
-----------------	----------------	----------

level	of transfer of content	of ST	of expression in TL	degree of task completion	mark
level 5	transfer of information; minor revision needed to reach professional standard.	Complete	Almost all the translation reads like a piece originally written in ST. There may be minor lexical, grammatical, or spelling errors.	Successful	,10
level 4	complete transfer; there may be one or two insignificant inaccuracies that require a certain amount of revision to reach professional standard.	Almost	Large sections read like a piece originally written in ST. There are a number of lexical, grammatical, or spelling errors.	Mostly successful	,8
level 3	of the general idea(s) but with a number of lapses in accuracy; Needs considerable revision to reach	Transfer	Certain parts read like a piece originally written in ST but others read like a translation. There	Adequate	,6

	professional standard.	are considerable number of lexical, grammatical, or spelling errors.	
		Almost	
	Transfer undermined by serious inaccuracies;	the entire text reads like a translation; there	
level 2	thorough revision required to reach professional standard.	are continual lexical, grammatical, or spelling errors.	adequate ,4
		The	
	Totally	candidate reveals a total lack of ability	T
level 1	of ST content; the translation is not worth revising.	to express himself/herself adequately in target language.	otally inadequate ,2

Finally, Waddington (2001, p. 315) defines the last method as, “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%”.

Sainz’s Rubric

Sainz (1992) introduces a student-centered approach to assessment of translations. She believes that teachers must make it clear that there are no right or wrong translations and that the students' translations are going to be used only as feedback for discussion later on. The process which she suggests for correction of translations comprises of five stages:

1. **Development:** a stage to understand and anticipate students' needs in order to satisfy those needs more efficiently.
2. **Implementation:** a stage during which students get the 'correction chart' shown in Table 6:

Table 6 – Sainz's rubric based on student-centered approach

M	Possible	S	Type of
Mistakes	Correction	Source	Mistake

Under the *Mistakes* columns, students write the word, phrase, or sentence which was understood as incorrect in their translation. Under *Possible Correction*, they try to produce an 'error free' version. The source of the answer for students' correction is entered under the column *Source* as 'I', 'Peer', 'Dictionary', and 'Teacher'. The column *Type of Mistake*, filled in by students, can serve as a good exercise to help them recognize what types of mistakes they are making and consequently eliminate them.

3. **Monitoring:** a stage to monitor the process by teachers to make adjustments as the course unfolds, on the basis of the information they retrieve from the 'correction chart'.
4. **Integration:** a stage during which teachers can fill in their own chart of 'types of mistakes' for a particular translation piece.
5. **Self-monitoring:** a stage during which students can check their own progress in the course and, at the same time, become critical about their learning.

At the bottom of the 'correction chart', students are asked to circle the figure, ranging from +3 to -3, which they think best matches their idea about their

performance on that particular translation passage and to make any other comments. A student-centered correction of translation is very useful in translation classes. By this careful system, the students' translations are subject to constant revision and changes in order to be improved. Small changes can sometimes create great effects.

Beeby's Rubric

Beeby (2000, p. 189) proposed a model for evaluation experienced in 1996 and 1997. In her model, the exam was marked out of 20:

- 10 points given for 10 specific translation problems that had been selected from the text
- 10 points given for language, and marks were subtracted for grammar mistakes: 1 point for incorrect syntax, tense, agreement and word order; ½ a point for incorrect articles, prepositions, and spelling

The 10 points selected for measuring translation competence were: point 1 for the translation of the headline and a full mark for an eye-catching headline; point 2 for typographical differences, for example capital letters in English and not in other languages; point 3 to 6 for the translation of three long sentences which involved transfer competence necessary in changing the tense, discourse competence necessary for splitting the sentence up and adding the information to different parts of the text to fit in with coherence expectations in English, and knowledge of syntactic differences between the two languages and understanding the function of the parenthesis in the ST; point 7 for relevance based on transfer competence, awareness of the TL readers' knowledge and making implicit information explicit where necessary; point 8 for lexical errors; point 9 for cultural transfer; and point 10 for extra-linguistic knowledge (Beeby 2000, pp. 192-5). The mentioned rubric is presented in Table 7.

Table 7 – Beeby's competence based rubric

Tr	
translation	headlines
elements	typography
	transfer
	discourse
	syntactic
	relevance
	lexical errors
	cultural transfer
	extra-linguistic
Points	

Goff-Kfourri's Rubric

Another rubric used to assess translation is proposed by Goff-Kfourri (2005) which declares that there are basically three options a rater can choose from when correcting a translation:

- 1 **General Impression:** Although some experienced raters are able to differentiate between a paper that is, for example, a 62/100 rather than a 67, a general impression mark is not very beneficial to the student for it does not provide the reasons for the missing marks.
- 2 **Error Count:** A simple error count is not recommended as a method of marking a student's translation since it rarely gives points for content and does not take into consideration the seriousness of the errors.
- 3 **Analytical Grid:** Heaton (1990, p. 110) proposed an analytical grid for language courses. However, it can be easily adopted for a translation assessment too. An analytical grid allows the instructor to set clear criteria for correction based on simple arithmetic (Table 8). In this particular case, the translation would be marked over 23 since the author chose to weigh mechanics less than the other areas.

Table 8 – Goff-Kfourri's rubric
Correction

Criteria

Fluency /Flow

Grammar
Terminology
General
Content
Mechanics

When students are provided with a grid assessment, they are able to see where their weaknesses and strengths lie. Some raters provide their students with a complete description of each number used on the grid. The objective of designing rubrics in education in general, and in translation specifically – either those based on errors or those that were more student-centered – has been their contribution and facilitation of different types of assessment. This is particularly true when it comes to assessing the more open-ended questions, ascertaining difficulties that might occur in the learning process.

Method

Subjects

Two groups participated in the two phases of this study. Both groups were instructors who were experienced in teaching and testing translation courses at different universities in Tehran. Five such instructors participated in the first phase of the study and completed the first questionnaire, while a total number of 50 instructors received the second questionnaire; 36 completed it.

Instrumentation

Two questionnaires were used in this study (see appendix). The first questionnaire, which was an open-ended one, was designed on the basis of reviewing several studies on translation issues, the pre-established rubrics and criteria for translation assessment, and interviewing translation students in universities about the tests and criteria for classroom translation assessment. The second one, which was structured and consisted of 22

Likert type items, was constructed on the basis of participants' answers to the open-ended questionnaire.

Procedure

As the first step, the open-ended questionnaire was developed as explained in the instrumentation section. After administering this questionnaire to five instructors of translation courses and receiving their responses, the second Likert type questionnaire was developed with 22 items. This questionnaire was then administered to 50 other translation instructors at different universities of Tehran. To elicit as accurate and objective responses as possible from the participants, the following points were taken into consideration: 1) Both the significance and the purpose of the study were explained to them, and 2) The participants answered the questionnaire items anonymously, so they felt at ease in answering the questions objectively.

Results

The results of this study which were based on an analysis of the responses provided by the instructors to the two questionnaires indicated that there was a general agreement among translation instructors in assessing students' translations. Almost all of the instructors (30 out of 36) preferred to use essay-type questions in translation exams, which is in accordance with the worldwide tendency. They believe that by this test type, the students may get better results in their translations, qualitatively and quantitatively. They also indicated their attention to accuracy (31 instructors), word equivalence (25), genre and rendering according to TL culture (23), register (21), grammatical points and reservation of style (19), and shifts (14) to assess students' translations.

According to Williams (2001), some translation researchers and theorists believe that assessment of translation is a subjective task but most of the participants of this study advocated the possibility of objective assessment of students' translations. This opinion held by Iranian instructors may be attributed to the fact that they suggest typical equivalents and translations for words and texts in the class and they expect students to emulate the same

translations at the time of exam. This might be what leads them to believe that students' translations can be assessed objectively.

Ninety-four percent of the instructors in this study (34 out of 36) agreed that they should inform their students of the intended evaluation criteria. They believe in explaining the criteria to their students before exams since it is one of the factors strengthening tests and helping students' answers to be directed to the required points.

There was an agreement among translation instructors about those criteria which are mostly ranked as primarily important for assessing the translations. Although some of the instructors were not familiar with translation theories, they intuitively applied the theories to assess students' translations. Thus, it can be claimed that the criteria considered to assess students' translations are teacher-made because most of the instructors chose them based on their experience in this field and not based on their knowledge.

The descriptive statistics of the responses provided on part A and part B of the second questionnaire are presented in Tables 9 below (To check the content of each item on the questionnaire, please see the appendix).

**Table 9 – Descriptive statistics for parts A and B of the second questionnaire (14 items)
Frequency and Percentage**

Item	Always +	Often +	Never
	Usually	Sometimes	r
	19 (52.8%)	17 (47.2%)	0
	25 (69.4%)	11 (30.6%)	0
	19 (52.8%)	17 (47.2%)	0
	31 (86.1%)	5 (13.9%)	0
	9 (25%)	22 (61.1%)	5 (13.9%)

	9 (25%)	18 (50%)	9 (25%)
	21 (58.4%)	9 (30.5%)	3 (8.3%)
	23 (63.9%)	11 (30.5%)	1 (2.8%)
	8 (22.2%)	12 (33.3%)	16 (44.4%)
0	23 (63.9%)	13 (36.1%)	0
1	13 (36.1%)	19 (52.8%)	4 (11.1%)
2	13 (36.1%)	21 (58.3%)	2 (5.6%)
3	14 (38.9%)	19 (52.7%)	0
4	17 (47.2%)	16 (44.5%)	3 (8.3%)

The descriptive statistics of the responses provided on part C of the second questionnaire are presented in Tables 10 below (To check the content of each item on the questionnaire, please see the appendix).

Table 10 – Descriptive statistics of part C of the questionnaire (8 items)

Frequency and Percentage

Item	S. Agree + Agree	Disagree + S. Disagree	Undecided
5	34 (94.4%)	0	2 (5.6%)
6	28 (77.8%)	4 (11.1%)	4 (11.1%)
7	13 (36.1%)	12 (33.3)	11 (30.6%)
8	30 (83.3%)	2 (5.6%)	4 (11.1%)
9	11 (30.6%)	14 (38.8%)	11 (30.6%)
10	9 (25%)	15 (41.7%)	12 (33.3%)
11	7 (19.5%)	21 (58.3%)	8 (22.2%)
12	28 (77.7%)	7 (19.4%)	1 (2.8%)

To find out whether the observed patterns of responses were significantly different from what we might have expected by chance alone, the Chi-square test of significance was run on the data. Table 11 presents the results of the Chi-square and as can be seen, the significant value for Chi for the items 6,

9, and 13 was not lower than 0.05 (0.793, 0.303, and 0.065, respectively), thus showing that the choice of the responses for these items was made by chance and they were, therefore, excluded from the findings.

Table 11 – Chi-square results for the frequency of the items

C				Ch			
tem	f	hi-square	ig.	tem	f	i-square	ig.
		7				23.	
		.575	056	2		917	000
		3				7.2	
		8.582	000	3		11	065
		1				24.	
		4.439	002	4		750	000
		9				27.	
		1.381	000	5		350	000
		9				76.	
		.243	055	6		410	000
		1				34.	
		.688	793	7		787	000
		4				58.	
		7.200	000	8		513	000
		5				21.	
		1.693	000	9		535	000
		4				23.	

	.847	303	0	639	000
		1		22.	
0	7.388	001	1	556	000
		1		15	
1	1.479	022	2	4.958	000

Finally and in accordance with the findings of the previously presented tables, the researchers proposed a detailed component-centered rubric in which the score of each component of the rubric is determined according to the percentage of its prevalence in the tables. Prior to that rubric though, the following table shows the percentages of each factor agreed by the instructors who participated in this study.

Table 12 – Percentages of factors decided to be used in the rubric

Items	S
	core
Accuracy	3
	0%
Finding the right and suitable word	2
equivalence in T.T	5%
TT's genre, TL culture	2
	0%
Grammar and preservation of style	1
	5%
Shifts	8
	%

The following table is the comprehensive and detailed rubric proposed by the researchers based on the findings of the study. The total score devoted to this rubric is 100. This rubric is a combination of the existing rubrics and the results and answers to the questionnaires used in this study. In making this rubric, different aspects of translation including comprehension and conveyance of sense and style among other factors have been taken into account.

Table 13 – The translation assessment rubric

Score range	Description
Accuracy (30%)	
25-30	No identifiable problems of comprehension; original message has been conveyed completely to TL readers; no omissions or additions to information
21-24	Virtually no problems of comprehension except with the most highly specialized vocabulary with no influence on TL readers' understanding; some partial omissions and additions
16-20	Information is conveyed to TL readers with some difficulty due to translator misunderstanding of some parts of original message; apparent omissions and additions
11-15	Poor expression of ideas; numerous serious problems in understanding ST interfere with communication of original message; difficult to understand TT
1-10	Severe problems interfere greatly with communication of original message; TL reader can't understand what original writer was trying to say
Finding equivalent (25%)	

20-25	All lexical and syntactic elements have been understood; precise vocabulary usage; words have been chosen so skillfully that the work reads like a good publishable version
15-19	Full comprehension and good usage of a wide range of vocabulary and structures; specialized vocabulary presents some problems with unsuitable equivalents
10-14	General comprehension of a fair range of vocabulary although some gaps observed; some vocabulary misused; some evidence of plausible attempts to work around difficulties of finding equivalents, perception, wordplay and other linguistic features
5-9	Comprehension of vocabulary and structures show quite noticeable gaps which obscure sense; problems in finding correct vocabularies; unable to cope with specialized vocabulary
1-4	Inappropriate use of vocabularies; comprehension of original seriously impeded even with fairly everyday vocabulary and structures; translation as a whole makes little sense
Register, TL culture (20%)	
17-20	Good sensitivity to nuances of meaning, register are precisely and sensitively captured; there is a sophisticated awareness of the cultural context; translation shows a sophisticated command of TL lexis, syntax, and register
13-16	There is a fair degree of sensitivity to nuances of meaning, register, and cultural context
9-12	There is a lack of sustained attention to nuances of meaning, register, and cultural context; no awareness of register; TL lexis, syntax, and register are not always appropriate
4-8	There is scant attention to nuances of meaning, register, and cultural context; there are serious to severe shortcomings in the use of appropriate lexis, syntax, and register
1-3	There is no appreciable understanding of nuances of

	meaning, register, and cultural context; no concept of register or sentence variety
Grammar and ST style (15%)	
13-15	Gives the feeling that the translation needs no improvement from grammatical and stylistic points though one or two natural failings might be observed; native-like fluency in grammar
10-12	Shows flair for stylistic manipulation of TL items as if text were written in TL originally except where the language is placed under severe pressure of comprehension; maintains advanced proficiency in grammar; some grammatical problems but with no influence on message
7-9	Tends to have awkward grammatical usage in TL and literalness of rendering though but not impeding sense in a significant manner; some attempts to reflect stylistic features of the original; some grammatical problems are apparent and have negative effects on communication
4-6	Clumsy TL; often nonsensical grammatical usages in TL; unnatural sounding; little attempt to reflect stylistic features of the original; there is evidence of clear difficulties in following style; grammatical review of some areas is clearly needed
1-3	Little sense of style which often makes poor sense in TL; knowledge of grammar is inadequate; use of TL grammar is inadequate; severe grammatical problems interfere greatly with message
Shifts, omissions, additions and inventing equivalents (10%)	
9-10	Correct use of relative clauses, verb forms; use of parallel structure; creative inventions and skillful solutions to equivalents; no fragment or run-on sentence
7-8	Almost all shifts appear with partial trespass, attempts variety; some inventions for not available equivalents in TL; no fragment or run-on sentence

5-6	Some shifts but not consistency; awkward and odd structure; only few run-on sentences or fragments present
3-4	Lacks variety of structure due to not preserving necessary shifts except for few cases; little or no evidence of invention in equivalents
1-2	Unintelligible sentence structure due to completely ignoring necessary shifts; no skillful handling of equivalents; no trace of invention

Conclusion

This study sought to develop an empirical rubric for translation quality assessment based on objective parameters of textual typology, formal correspondence, thematic coherence, reference cohesion, pragmatic equivalence, and lexico-syntactic properties.

As different rubrics and criteria for translation assessment and also the results of this study among translation instructors show, the emphasis is laid on the importance of objective and standard assessment in the field of translation and translation teaching.

The findings of this research and the rubric presented can serve translation instructors in order to come up with a more objective assessment of students' translation works. Students majoring in translation can also benefit from the findings of this study too since they would certainly be able to improve their translations if they were aware of the comprehensive criteria used to evaluate their translations.

The Authors

Hajar Khanmohammad is Assistant Professor of Applied Linguistics at Islamic Azad University, Central Tehran Branch. She has published several articles in academic journals and conference proceedings. She has also translated books and articles from English into Farsi. Dr. Khanmohammad's main areas of research include testing and assessment and also translation studies.

Maryam Osanloo received her MA in translation studies from Islamic Azad University, Central Tehran Branch, in 2008. She has experience working both as an English teacher and a professional translator.

References

- Al Qinaï, J. (2000). *Translation quality assessment, strategies, parameters, and procedures*. Retrieved on February 10, 2009 from:
<http://www.erudit.org/revue/meta>
- Bassnett, M. S. (1980). *Translation Studies*, London and New York: Routledge.
- Beeby, A. (2000). *Teaching translation from Spanish to English*. Ottawa: University of Ottawa Press.
- Bell, R. T. (1991). *About translation*. London: Longman
- Belloc, H. (1931). *On translation*. Oxford: Oxford University Press.
- Catford, J. C. (1965). *A linguistic theory of translation*. London: Oxford University Press.
- Farahzad, F. (1992). Testing achievement in translation classes. In C. Dollerup & A. Loddergard (Eds.), *Teaching translation and interpreting* (pp. 271-278). Amsterdam/Philadelphia: John Benjamins.
- Gerasimov, A. (2005). *My golden rules for quality assurance*. Retrieved on March 20, 2009 from:
<http://fra.proz.com/doc/534>
- Goff-Kfourî, C. A. (2005). Testing and evaluation in the translation classroom. *Translation Journal*, 9(2), 75-99.
- Hatim, B., & Mason, I. (1997). *Translator as communicator*. London: Routledge.
- Heaton, G. B. (1990). *Writing English language tests*. London: Longman.
- Mason, I. A. (1997). Pragmatic computation. In J. Harland (Ed.), *Proceedings of the Australian Theory Symposium Cats' 97* (pp. 103-112). Retrieved on Feb 20, 2009, from:
<http://mcs.une.edu.au/~iam/Data/papers/97cats.ps>
- McTighe, J. (1997). What happens between assessments? *Educational Leadership*, 6(12), 110-136.
- Muzii, L. (2007). Adaptation, translation, and multimediality. *Online Translation Journal*, 1(4), 46-58.
- Nida, E. A. (1984). *Translation theory and practice*. India: UBS.
- Rico Perez, C. (2002). Translation and project management. *Translation Journal*, 6(4), 38-52.
- Riazi, A. M. (2003). The invisible in translation: The role of text structure in translation. *Journal of Translation*, 7, 1-8.
- Sager, J. C. (1983). Quality and standards: The evaluation of translations. In C. Picken (Ed.), *The translator's handbook* (pp. 91-102). London: ASLIB.
- Sainz, J. M. (1992). *Student-centered correction of translation*. Amsterdam/Philadelphia: John Benjamins.
- Savory, T. H. (1968). *The Art of translation*. Oxford: Pergamon Press.

- 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 on February 10, 2009 from:
<http://catalog.ebay.ca/introduction-to-rubrics>
- Tajvidi, G. R. (2005). Translation quality assessment. *Translation Studies*, 10, 27-40.
- Virginia Education Association and Appalachian Educational Laboratory (1992). Recommendations for teachers in authentic assessment resources. Retrieved on Feb 10, 2009 from:
<http://people.ucsc.edu/~ktellez/authenres.html>
- Waddington, C. (2001). Different methods of evaluating student translation: The question of validity. *Met: Translator's Journal*, 46(2), 311-325.
- Wangsatorntankhun, J. A. (2007). Designing performance assessment challenges for the three-story intellect. Retrieved on Oct 30, 2008 from:
www.Geocities.com/Athens/Parthenon/8658
- Williams, C. (2001). Language and education in a non-territorial developing country. *EID Review*, 5(11), 1462-1614.

Questionnaires

The open-ended questions:

1. How do you evaluate students' translation performances?
2. What factors do you consider when evaluating students' translations?

The structured questionnaire:

Part A: Following are factors your colleagues consider in evaluating students' translations. Please mark the frequency of the factors as they apply to you when grading students' translations.

1. I consider Grammar as a criterion in my evaluation.
Always Usually Often Sometimes Never
2. I look for appropriate use of Word Equivalents as a criterion in my evaluation.
Always Usually Often Sometimes Never
3. I attend to Style as a criterion in my evaluation.
Always Usually Often Sometimes Never
4. I check Accuracy in rendering the exact message from the source language in my evaluation.
Always Usually Often Sometimes Never
5. I consider Beauty as a criterion that should be kept in my evaluation.
Always Usually Often Sometimes Never
6. The more similarity between students' translations and my own translation, the better mark students will obtain.
Always Usually Often Sometimes Never

Part B: The following items ask about the things students are required or allowed to do in their translations. Please choose the most appropriate option.

7. Students must take care of the Registers in their translations.

- Always Usually Often Sometimes Never
8. Students must take care of the Genres through the translation.
Always Usually Often Sometimes Never
9. Students must translate the text exactly according to the SL culture even though it may not make sense in the TL.
Always Usually Often Sometimes Never
10. Students must observe the differences between TL & SL cultures and translate the text according to the TL culture.
Always Usually Often Sometimes Never
11. Students have the chance to invent equivalent for the words that have no proper equivalent in the TL.
Always Usually Often Sometimes Never
12. Additions and Omissions on the part of students are acceptable in translation.
Always Usually Often Sometimes Never
13. Catford's* Shifts are acceptable in translation on the part of the students.
Always Usually Often Sometimes Never
14. I think objective assessment of translation is possible.
Always Usually Often Sometimes Never

Part C: The following are other instructors' attitudes toward different aspects of students' translations. Please select the choice you keep in mind in preparing translation tests.

15. Students should be informed of my evaluation criteria before exam.
Always Usually Often Sometimes Never
16. Students are allowed to use dictionary in the exam.
Always Usually Often Sometimes Never
17. I think Multiple-Choice is a proper test type to assess students' translations.
Always Usually Often Sometimes Never
18. I think Essay type is a proper test type to assess students' translations.
Always Usually Often Sometimes Never
19. I think Completion Test is a proper test type to assess students' translations.
Always Usually Often Sometimes Never
20. I think Cloze Test is a proper test type to assess students' translations.
Always Usually Often Sometimes Never
21. I think T/F statement is a proper test type to assess students' translations.
Always Usually Often Sometimes Never
22. I think the test items should be contextualized (put in a context).
Always Usually Often Sometimes Never

