How to evaluate the TEFL students' translations: through analytic, holistic or combined method?

Mojtaba Amini

Correspondence:
m.amini257@yahoo.com
Department of English Language,
University of Isfahan, Isfahan, Iran

Abstract

Background: Translation quality assessment (TQA) suffers from subjectivity in both neighboring disciplines: 'TEFL' and 'Translation Studies, and more empirical studies are required to get closer to objectivity in this domain. The present study evaluated the quality of the written translation of TEFL students through three different approaches to translation quality assessment (TQA) in order to examine the efficiency and reliability of the three methods and ultimately suggest the most reliable one.

Methods: Thirty BA TEFL university students translated a text from English into Persian, and three raters scored the translated texts through three different methods of assessment.

Results: The results of statistical analysis indicated that, the error analysis method B was more reliable than holistic method C, but was less reliable than combined method D.

Conclusions: That is, when we combined the results of both error analysis and holistic methods in a proportion of 70/30, the new combined error analysis-holistic method got a better reliability rating, and accurate results than holistic and analytic methods. Therefore, the combined method may be suggested as a reliable method for evaluating and scoring the TEFL students' translations.

Keywords: TEFL students, Translation quality assessment (TQA), Analytic method, Holistic method, Combined method

Background

Assessment and evaluation play an important role in teaching English as a foreign language (TEFL), because based on the reliable assessment, one can talk about the degree of the efficiency of the teaching methods and improvement of the language learners. In effective instructional programs, assessment provides convincing evidence to the participants that the curriculum goals and objectives are being met. As Sawyer (2004) put, high quality education is based upon sound assessment. There may be various circumstances for assessment; Sager (1989) distinguishes between two main and different settings where evaluation occur: evaluation of professional and academic translation, and Williams (1989) believes in difference between assessment in an academic environment and assessment in the professional atmosphere as well.

Translation quality assessment (TQA) suffers from subjectivity in both neighboring disciplines: "TEFL" and "translation studies." Al-qinai (2000) states that over the last two decade, studies (House 1977; Wilss 1982; Hatim & Mason 1990, Baker 1992; Horton 1998) have ventured to introduce objectivity instead of subjective impressionism in judging translation quality. Although different methods have been proposed to decrease the subjectivity in translation evaluation and to get closer to objectivity and reliability, their findings have not satisfied the experts of the field, and this domain calls for further studies.

Quality in translation is hard to define and may be regarded as controversial to some extent. House (1997) states that evaluating the quality of a translation presupposes a theory of translation. Therefore, different views of translation lead to different concepts of quality in translation and hence different ways of assessing it. Koponen (2010) claims the concept of quality covers accuracy, fluency, and fitness for the purpose, while Mateo (2014) states that the search for quality in translation is still an unsolved issue today. From the second half of the twentieth century onwards, there have been controversial issues regarding the quality concept and the way to determine it. It seems that there is no common ground when it comes to defining quality in translation either from a practical or from a theoretical viewpoint. As Colina (2009) put, it has been the excess of conflicting opinions and the experts' lack of consensus on the definition of quality that have hampered any potential progress in the field, and many scholars still believe that quality in translation is a relative and subjective concept (Horguelin & Brunette 1998; Larose 1998; Parra 2005).

Translation studies suggest that the trends of quality assessment can be broadly grouped into two categories: those based on error analysis and those based on a holistic approach, with some attempts to combine the two. That is, the common TQA methods which are applied in academic settings include holistic, analytic, and combined methods. In addition to these three approaches, the corpus-based methods have been introduced recently, and they have turned to be efficient as well. McAlester (2000) argues that in study of assessment in translation, it is naturally desirable that the methods used for assessment be reliable, valid, objective, and practical, while Williams (2004) believes that the most contentious issue in TQA is the lack of uniformity in assessment of language errors, and Sager (1983) even goes beyond the errors and proposes that translation assessment should take into account not only different types of error but also the effect of a particular error on the whole text.

The reported results on the reliability and efficiency as well as the merits and demerits of different methods of TQA are mixed but are slightly in favor of the analytic methods. Some investigations have concluded the analytic methods as more efficient and reliable while others have reported holistic methods as more reliable compared to analytic. As for the combined methods, they have been almost untouched in empirical studies. Further, as Bahameed (2014) put, the empirical studies examining the reliability of TQA methods have been relatively few in number and have been touched mildly, some of which include (Waddington 2001, 2003; Garant, 2009; Zamani Delshad, 2012; Bahameed, 2014; Phelan 2017).

Waddington (2001) studied different methods of evaluating student translations, during which supported the criterion-related validity of all four systems of assessment (methods A, B, C, and D). In his probe to introduce the most reliable method of assessment, Waddington (2003) applied methods A, B, C, and D for scoring Spanish-English translations and concluded that the two error analysis methods A and B are equally reliable and more reliable than holistic method C. However, method D which is the

combination of methods B and C in a proportion of 70/30 is the most reliable method. Unlike Waddington, Garant (2009) during a case study reported the general trend toward holistic method as a significant finding. In line with Waddington, Zamani Delshad (2012) concluded that the error analysis method was more accurate and reliable than the holistic one; she made comparison between error analysis and holistic approaches in translation quality assessment of journalistic texts from English into Persian. Further, she reported the holistic method as more practical which needed lesser time than error analysis to be applied. Bahameed (2014) applied Hurtado's method of evaluation (which is based on error analysis) on female translators and supported the reliability and efficiency of the error analysis method. The method was applied to the correction of 43 female students' translations of the final exam containing different texts to be translated in both directions between English and Arabic. The method was found out to be reasonable to give impartial translation quality evaluation for the students' translations. Unlike Bahameed (2014), Phelan (2017) examined one of the analytic methods and reported some demerits of it. Phelan (2017) during a case study applied the American Translators Association framework (ATA) to assessment of legal translations; ten raters scored the translations based on the framework. Raters' feedback indicated that some error categories overlapped or were vague and the flowchart was difficult to implement, in particular when deciding the level of seriousness of errors. She concluded that at first sight, the ATA framework gave the impression of being an analytical approach with very little possibility of subjectivity on the part of the assessors, but it turned out to be quite subjective when implemented. Waddington (2001) sent out a questionnaire to 48 European and Canadian universities to examine the degree of application of various methods of TQA by teachers. A total of 52 teachers replied from 20 of these universities; the result was as follow: 36.5% of the teachers used a method based on error analysis, 38.5% used a holistic method, and 23% used a combined error analysis with a holistic appreciation. Based on his study, holistic methods are more common than the other types.

The small number of existing empirical studies in this domain as well as mixed reported results calls for further investigations. In addition, studies with focus on English-Persian language pair examining the efficiency of the combined methods of translation evaluation were not found in the literature. Thus, the present empirical study intended to fill the gap in this domain by focusing on English-Persian language pair which has been almost untouched. Hence, the error analysis method B, the holistic method C, and the combined method D were examined to answer the following question:

> "Can the quality of English-Persian translation be assessed more accurately if the method of assessment combines error analysis with a holistic appreciation?"

Method B, C, and D
In this section, methods B, C, and D are introduced briefly:

Method B
Waddington' method (Waddington 1999) which is called method B (See table 1), examines the penalty for each mistake according to the extent of its effect on the overall quality of the concerned translation. The corrector first has to decide 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: if it does not, it is a language mistake (and is penalized with − 1 point), if it does, it is a translation mistake (and is penalized with − 2 points). However, in the case of translation mistakes, the corrector has to judge the importance of the negative effect that each one of these mistakes has on the translation, taking into account the objective and the target reader specified in the instructions to the translator for each translation. In order to judge this importance, the corrector is given in Table 1.

Method C

Method C is a holistic method of correction with the following features (See Table 2):

1. It presents a unitary scale which considers the translation competence as a whole, instead of dividing it into various sub-scales representing different sub-competences.
2. The descriptors do not use terminology that would presuppose specialist knowledge (such as applied linguistics) on the part of the correctors.
3. It includes only five main levels in an attempt to achieve maximum consistency between raters (see Pollitt 1991:90), although there are two marks within each level in line with the traditional Spanish system of marking (from 0 to 10).

Method D

Method D is the combination of error analysis method B and holistic method C in a proportion of 70/30, that is, method B accounts for 70% of the total result and method C for the remaining 30%.

Method

Participants

Thirty BA EFL university students, both males and females and without any age limitation, participated in the study and translated an English text into Persian.

Materials

Instruments of the study included an English text (with general content) including 266 words as a translation task(see Additional file 1: Appendix S1); students had 2 h to translate it. Other instruments were laptop computer and SPSS software for data analysis.

Table 1 Marking system and typology of errors in method B (Waddington 1999)

Negative effect on words in ST	Penalty for negative effect
On: 01–05 words	2
06–20 words	3
21–40 words	4
41–60 words	5
61–80 words	6
81–100 words	7
100+ words	8
The whole text	12

Table 2 Holistic method C

Level	Accuracy of transfer of source text content	Quality of expression in TL	Degree of task completion	Mark
5	Complete transfer of ST information; only minor revision needed to reach professional standard.	Almost all the translation reads like a piece originally written in English. There may be minor lexical, grammatical, or spelling errors.	Successful	9, 10
4	Almost complete transfer; there may be one or two insignificant inaccuracies; requires certain amount of revision to reach professional standard.	Large sections read like a piece originally written in English. There are a number of lexical, grammatical, or spelling errors.	Almost completely successful	7, 8
3	Transfer of the general idea(s) but with a number lapses in accuracy; needs considerable revision to reach professional standard.	Certain parts read like a piece originally written in English but others read like a translation. There are a considerable number of lexical, grammatical, or spelling errors.	Adequate	5, 6
2	Transfer undermined by serious inaccuracies; thorough revision required to reach professional standard	Almost the entire text reads like a translation; there are continual lexical, grammatical, or spelling errors.	Inadequate	3, 4
1	Totally inadequate transfer of ST content; the translation is not worth revising.	The candidate reveals a total lack of ability to express himself adequately in English.	Totally inadequate	1, 2

Procedure

Thirty BA EFL students translated an English text into Persian then the translations were collected for scoring and evaluating (see a sample of the translated Persian text in Additional file 1: appendix 2). In addition to the researcher, two other raters (both PhD candidates in translation studies) were trained regarding the application of these three methods of translation assessment. Raters studied each method individually and in different orders (method B/C/D, D/C/B, C/D/B) to control the variable of one method influencing the following one. There was also an interval of 3 weeks between the application of one method and the next. After finishing the scoring procedure, data was analyzed through SPSS software.

Results

The inter-reliability

The results of inter-rater reliability indicated that error analysis method B (.93) was more reliable than holistic method C (.88) and was less reliable than method D (.97) (see Table 3).

ANOVA and post-ANOVA

The discrepancy between marks turned out to be so mush. Therefore, through variance analysis, the source of this variance could be determined: was this variance caused by the students (the rows) or by the correctors applying the method (the columns)? One may

Table 3 Results of inter-rater reliability

Method	Inter-rater reliability
Method B	.93
Method C	.88
Method D	.97

claim the difference of the results of the three methods can be attributed to various students' translation competence, because translations may cover the wide range from excellent to inadequate, and one would hope that the variance detected would be mainly explained by these differences, and less variance would be caused by the correctors applying the methods. The analysis of the variance (ANOVA) was primarily consistent with the results of inter-rater reliability, but it revealed significant amounts of variance in both rows (the students) and columns (the correctors applying the methods).

To calculate the size of the two sources of variance, post-ANOVA analysis called eta-squared was applied, which gave the results shown in Table 4. The analysis indicated that with method B, *55%* of the variance detected can be attributed to the rows, i.e., the difference between the students, whereas only *12%* to the columns, i.e., the raters applying the methods. These values were better compared to those of method C: *34%* as against *19%*., but with the combined method D, we achieved the best results of all: *76%* as against *5%*.

Discussion

The present study intended to answer the following question through examining three methods of TQA and to suggest the most efficient and reliable one:

> "Can the quality of English-Persian translation be assessed more accurately if the method of assessment combines error analysis with a holistic appreciation?"

Based on the results of the statistical analysis, it was found that error analysis method B was more reliable than holistic method C and less reliable than combined method D. That is, combined method D turned out to be the most reliable of the three methods.

Considering the results approving the higher reliability of the analytic method compared to the holistic one, the present study was in line with some studies in this domain, including (a) Waddington (2003) who examined methods A, B, C, and D in Spanish-English language pair and reported analytic methods as more reliable than holistic one, the main difference was that he applied two error analysis methods instead of one; (b) Zamani Delshad (2012) who made comparison between error analysis and holistic approaches in translation quality assessment of journalistic texts from English into Persian and reported error analysis as more reliable than holistic one, the difference was that Zamani Delshad did not include the combined method in her study; and (c) Bahameed (2014) who applied Hurtado's error analysis method and reported the method as reliable and efficient.

It seems that in analytic methods, the precision and rigidity of various subcomponents refrain the raters from deviation from the principles of the rubric, and such precision and detailed scrutiny leads to the efficiency of the method.

Table 4 Coefficients of eta-squared for methods B, C, and D

			η^2
Method B	η^2 rows =	$\frac{340}{623.43}$	.545
	η^2 columns =	$\frac{73.5}{623.43}$	.117
Method C	η^2 rows =	$\frac{240}{714}$	.336
	η^2 columns =	$\frac{132.5}{714}$	.185
Method D	η^2 rows =	$\frac{380}{503.23}$	.755
	η^2 columns =	$\frac{28}{503.23}$	.05

The present study approved the combined method D as the most reliable method among the other ones which was consistent with Waddington (2003). This result allows us to implicate that when error analysis method B is combined with holistic method C in proportion of 70/30, the weaknesses of one method can be compensated with the strengths of the other one, which decreases the subjectivity in evaluation and leads to the highest reliability and efficiency of the combined method D compared to other methods.

The results of the three methods applied in the present study may be dependent upon the direction of the translation. That is, if the direction of translation was from native to the foreign language, the results might possibly be different, because translating into a foreign language can be more susceptible to making error than translating from a foreign language. Furthermore, text type may have a role in the outcome of the study, that is, different text types may be evaluated more efficiently by applying certain TQA method rather than the other one which needs to be investigated.

Conclusion

Considering the mixed results of the previous empirical investigations on the analytic and holistic methods of TQA and lack of enough studies dedicated to the combined methods, the present study examined the reliability of three methods of TQA in English-Persian translation. Statistical analysis including inter-rater reliability, ANOVA, and post-ANOVA results supported the higher reliability of the analytic method than the holistic one and also the highest reliability of the combined method.

The application of combined methods for scoring and evaluating translation may increase the reliability of the TQA outcome in academic settings. Therefore, teachers and professors can benefit more from combined methods to assure the reliability in their scoring procedure. However, further empirical studies are required in order to enable us to talk with certainty, because in different circumstances, various methods may turn out to be helpful.

The result of the present study was primarily limited to the number of the participants and consequently to the number of translations. As was mentioned earlier, 30 translations were applied for analysis. Future studies can be carried out with more participants who can provide the study with more data for analysis. Among the other limitations of the study, it can be referred to the number of the raters. It is suggested that other studies benefit from more raters if possible to come to a more reliable outcome. It is also suggested that future studies be carried out by (a) changing the direction of translation (from Persian into English) and (b) focusing on other language pairs which have not been investigated so far in this domain.

Additional file

Additional file 1: Appendix S1. Sample of the original text in English. Appendix S2. Sample of the translated text into Persian. (DOCX 32 kb)

Authors' contributions
The author confirms being the sole contributor of this work and approved it for publication.

Authors' information
Mojtaba Amini is the PhD candidate in English Language-Translation Studies at the University of Isfahan, Isfahan, Iran. Cognitive approaches to interpreting, ideology and translation, and translation quality assessment (TQA) are among his fields of interest.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Al-Qinai, J. (2000). Translation Quality Assessment. Strategies, Parameters and Procedures. *Translators' Journal*, *45*(3), 497–519.
Bahameed, AS. (2014). Evaluating Female Translators. *International Journal of Comparative Literature & Translation Studies*, *2*(2), 59–65.
Baker, M (1992). *In Other Words. A Coursebook on Translation*. London: Routledge.
Colina, S. (2009). Further Evidence for a Functionalist Approach to Translation Quality Evaluation. *Target*, *21*(2), 215–244.
Garant, M. (2009). A case for holistic translation assessment. *AFinLA-e Soveltavan kielitieteen tutkimuksia*, *1*, 5–17.
Hatim, B, & Mason, I (1990). *Discourse and the Translator*. London/New York: Longman.
Horguelin, P, & Brunette, L (1998). *Practique de la Révisión, 3ème Edition Revueet Augmenteé*. Québec: Linguatech éditeur.
Horton, D. (1998). Translation assessment: Notes on the interlingual transfer of an advertising text. *IRAL*, *36*(2), 95–119.
House, J (1977). *Translation Quality Assessment*. TBL-Verlag Narr: Tubingen.
House, J (1997). *Translation quality assessment: a model revisited*. Tübingen: Gunter Narr Verlag.
Koponen, M. (2010). Assessing Machine Translation Quality with Error Analysis. In *Electronic proceedings of the KäTu symposium on translation and interpreting studies* 4 (pp. 1–12). (URL http://www.sktl.fi/@Bin/40701/Koponen_MikaEL2010.pdf).
Larose, R. (1998). Méthodologie de l'Évaluation des Traductions. [A Method for Assessing Translation Quality]. *Meta*, *43*, 163–186.
Mateo, RM. (2014). *A Deeper Look in Metrics For Translation Quality Assessment (TQA): A Case Study. miscelánea: a journal of English and American studies*, *49*, 73–94.
McAlester, G (2000). The evaluation of translation into a foreign language. In C Schaffner, B Adab (Eds.), *Developing Translation Competence, Benjamins Translation Library 38* (pp. 229–241). Amsterdam: Benjamins.
Parra, GS (2005). *La Revisión de Traducciones en la Traductología: Aproximación a la Práctica de la Revisión en el Ámbito Profesional Mediante el Estudio de Casos y Propuestas de Investigación, Doctoral Dissertation* (). Spain: Universidad de Granada.
Phelan, M. (2017). Analytical assessment of legal translation: a case study using the American Translators Association framework. *The Journal of Specialized Translation*, *27*, 189–210.
Pollitt, A (1991). Response to Charles Alderson's Paper: Bands and Scores. In C Alderson, B North (Eds.), *Language Testing in 1990s*, (pp. 87–94). London: Macmillan.
Sager, JC (1983). Quality and standards: The evaluation of translations. In C Picken (Ed.), *The translator`s handbook*, (pp. 91–102). London: ASLIB.
Sager, JC (1989). Quality and Standards – the Evaluation of Translations. In C Picken (Ed.), *The Translator's Handbook*, (pp. 91–102). London: ASLIB.
Sawyer, DB (2004). *Fundamental Aspects of Interpreter Education*. In *Amsterdam/Philadelphia: John Benjamins Publishing Company*.
Waddington, C (1999). *Estudio comparativo de diferentes métodos de evaluación de traducción general (Inglés-Español)*. Madrid: Publicaciones de la Universidad Pontificia Comillas.
Waddington, C. (2001). Should translations be assessed holistically or through error analysis? *Hermes, Journal of Linguistics*, *26*, 15–38.
Waddington, C (2003). A Positive Approach to the Assessment of Translation Errors. In MM Ricardo (Ed.), *I AIETI. Actas del I Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. Granada 12-14 de Febrero de 2003*, (pp. 409–426). AIETI: Granada.
Williams, M. (1989). The assessment of Professional Translation Quality: Creating Credibility out of Chaos. *TTR Traduction, Terminologie, Redaction*, *2*(2), 13–33.
Williams, M (2004). *Translation Quality Assessment: An Argumentation- Centered Approach*. Ottawa: University of Ottawa Press.
Wilss, W (1982). *The Science of Translation. Problems and Methods*. Tübingen: Gunter Narr.
Zamani Delshad, M. (2012).*1The Comparative Study of Error Analysis versus Holistic Approach in Translation Quality Assessment of Journalistic Texts*. Unpublished MA Thesis, IAU, Central Tehran Brach, Tehran.