

Investigating the Variations among Translation Assessors when Assessing the Translation Quality of Trainee Translators

Whaj Muneer Esmail¹ & Mahmoud Ali Ahmed² - Sudan University of Science and Technology, Faculty of Languages²

ABSTRACT:

This paper addresses a very important topic in the field of translation training and translation quality assessment from academic and professional perspectives. It aims at investigating the assessment variations among the translation assessors. Moreover, there are two objectives of this paper, namely; to identify the variations among translation assessors in assessing the translation quality of the trainee translators and to identify factors might affect the process of assessing the translation quality. The assessment of translation quality has not taken much luck in translation studies. Because of the importance of this field, that has many contradictions, and the existence of several divergent views on the assessment of translation quality, this research tackles two tests: written and simultaneous interpreting. Their purpose is to collect data for statistical processing by using SPSS statistics. Thereafter, a t-test was carried out to specify the differences in mean values of the assessment for each trainee translator. The results of the research revealed inevitable and obvious variations among translation assessors when assessing the translation quality of trainee translators/interpreters. This paper recommends further studies deeply goes into its important topic. It also recommends studying factors that might affect the assessment process of the translation quality.

Keywords: Translation Quality Assessment, Translation Assessors, Assessment Variations

المستخلص:

يتناول هذا البحث عنوانا مهما في مجالات تدريب الترجمة وجودتها من منظور اكايمي و مهني. لم تحظى جودة الترجمة بالأهتمام الكبير في الدراسات الترجمية. يهدف هذا البحث الى استقصاء التباين الحاصل بين المقيمين في تقييم جودة الترجمة حيث يكمن الغرض من هذا البحث في امرين: الاول تحديد التباين بين تقييمات المقيمين لجودة الترجمة للمترجمين من هم تحت التدريب والثاني تحديد العوامل التي قد تؤثر عملية تقييم جودة الترجمة للمترجمين. وبسبب اهمية هذا الميدان المليئ بالتناقضات ووجهات النظر المتباينة حيال تقييم جودة الترجمة اقتضى البحث اجراء اختبارين لجودة الترجمة (ترجمة تحريرية وترجمة فورية) الغرض منهما هو جمع بيانات عن التباين في تقييم جودة الترجمة ومن ثم تحليل هذه البيانات بأستعمال برامج (SPSS-23) لغرض تحليل البيانات احصائيا ومن ثم استعمال اختبار (T-test) لغرض حساب قيم الأوساط الحسابية (mean values) للتباين في تقييم جودة الترجمة. اذ اظهرت نتائج البحث بأن هناك تباين معنوي واضح جدا في تقييمات المقيمين لجودة الترجمة لعينة من المترجمين المتدربين. يوصي هذا البحث الى اجراء بحوث اخرى تتناول و بعمق مجال تقييم جودة الترجمة من حيث الحرص على اجراء بحوث تدور حول العوامل التي قد تؤثر على سير عملية تقييم جودة الترجمة.

الكلمات المفتاحية: تقييم جودة الترجمة - مقيم الترجمة - التباين في التقييم.



INTRODUCTION:

Considering a quality of any piece of translation is quite good or not, depends entirely on the individual receiving the translated text. Oxford English Dictionary (2015) defines the word quality as “*How good or bad something is*”. People may have different views regarding different issues; this interprets that the quality of something might be accepted for someone and not accepted by somebody else. In this regard, we deal with an inextricable matter that cannot be judged easily. As far as translation quality is concerned, it is the same with any quality, such as a color, design, or style. Therefore, the term *QUALITY* is a flavor belongs to a personal insight, and for this reason, it is considered as a subjective matter that differently digested by different people. Translation quality assessment per se is considered, as a subjective matter ought to be submitted to certain criteria. These criteria are controversial among translation scholars since there are no specific models adopted for the assessment of translation quality and this is exactly what the review of the literature indicates on the assessment of the translation quality (Hang, 2019). Most proposed criteria differ depending on the purpose of the assessment and the theoretical framework applied by those who involved in assessing the translation quality.

After wide observations, the researcher noticed that most assessors assess translators, students or trainee translators differently. Seemingly, they apply different criteria or do not depend on certain criteria. This paper tries to answer the question: “*What significant variations do translation assessors have when assessing the translation quality of trainee translators?*”

Accordingly, this paper hypothesizes: “*There are significant variations among translation assessors when assessing the translation quality of trainee translators.*”

This paper employs two translation tests *Written Translation and Simultaneous Interpreting*. These two tests were conducted in order to investigate the potential variations among translation assessors in the process of translation quality assessment (TQA). Statistical analysis was done with SPSS version-23 to the assessment sheets. The study also use one sample t-test in order to test the mean of the variations in the assessment sheets to find out whether it significantly different from the hypothesis.

Related Literature:

Innumerable studies have chased TQA in the educational or training environments. All of which could not reach typical solutions for handling translation quality (TQ) with a clear vision. This might be contributed to a variety of factors related to the assessors, assessment criteria or trainee translators, and so on. TQA as defined by (Bowker, 2000 p. 183) is a very complicated issue that needs lots of attention by the assessors. Furthermore, many other scholars have also noted that TQA is a very problematic area deserves deep meditations by translation researchers. Scholars, such as House 1977; Reiss et al., 2015; Newmark 2008, and others believe that there are many difficulties facing translation students, new translators, professional translators, teachers and translation assessors in dealing with the TQA, translation process or the field translation itself. Difficulties might be with the subjectivity assessor might practice during the assessment process of TQ.



Another problematic issue: there is no specific model available can fill the knowledge gap of applying certain criteria of TQA.

Moreover, most scholars also stress that the models at hand are full of ambiguity. These models cannot be applied in the educational or training contexts, as well. Some models such as the functionalist model has already been applied to analyze different typical of technical genres in translation, including commercial texts.

This model is hard to apply within the education context due its complexity in its application.

A study done by (Liu and Zhao, 2016, p. 558) regarding TQA model, the authors depended on Malcolm Williams' Argumentation-centered. They demonstrated this model and its parameters of assessing TQ. This model depends on the argumentative theory, which does not provide a full subjective assessment to TQ (Liu and Zhao, 2016, p 561).

Therefore, it is well known to every assessor that he cannot depend on a certain model in his assessment to TQ.

Briefly, Malcolm Williams's argumentative model of TQA is based on argumentative theory that focuses on a macrotexual approach showing whether the TQ is acceptable or not based on a linguistic background (Clarke, 2008). Moreover, the researcher point of view that this model is bound to the macrotexual level of translation assessment ignoring many issues such as the text type, the purpose, translation strategies, and so on. Accordingly, this model has a little chance of possibility to be applied in the field of training or assessment to trainee translators. House model (1977) of TQA based on Halliday's systemic functional

theory is also another pioneering functional-pragmatic model. She emphasizes the role of comparing the function of a source and target text. However, her model of TQA is very hard to be applied to training translators. Furthermore, her model of TQA has a great deal of criticisms by some observers in the field of translation. House's model of TQA lacks to name some factors might affect the assessment process of TQ.

Research Methodology

After noticing the variation of the assessors' assessment of TQ, the researcher decided to investigate if there is any variation translation assessors practice in the assessment process of TQ.

Based on that assumption, two tests were designed to achieve the purpose of this study. The first one is a written translation test in which a group of students participates to deliver their translation versions. The second test was a simultaneous interpreting that involves a group of participants to record their interpretations via two Mp-3 audios. Their recordings were collected and delivered to six assessors to assess the TQ of the answer sheets. The sample of the participants in these two tests was trainee translators and interpreters who recently graduated from departments of translation. The sample in question is a group of translation assessors who have long experience and relevant background in the field of translation for several years. Thereafter, descriptive statistics were done by SPSS version-23 was used for the purpose of analyzing the data of the written translation test and simultaneous test. T-test is another statistical tool was involved in this study. This tool measures the mean values of the variations derived from the descriptive statistics resulted from SPSS.

• Witten Translation Test

Two translation tests were conducted in order to be assessed by translation assessors. The first one was a media written text of about 600 words. The participants were asked to translate the English text into Arabic and the Arabic text into English. The participants were 15; however, only twelve took part in this test.

Thereafter, the answer sheets with the assessment sheets were sent to 12 assessors

in order to assess the translation quality of the 12 participants. Only ten assessors were responded. The scores out of 20 marks displayed in the assessment sheets were indexed into a table. The names of the translation assessors and the trainee translators were coded in numbers for the purpose of anonymity as shown below in table (1).

Table (1) Marks out of (20)
ASR = Assessors TT = Trainee Translator

Details	TT - 1	TT - 2	TT - 3	TT - 4	TT - 5	TT - 6	TT - 7	TT - 8	TT - 9	TT - 10	TT - 11	TT - 12
ASR - 1	17	11	16	15	8	9	13	18	8	11	16	7
ASR -2	12	16	12	12	9	10	7	12	11	10	12	12
ASR -3	16	17	18	19	15	15	12	14	16	17	14	17
ASR -4	13	12	8	15	10	7	8	11	7	9	11	15
ASR -5	10	14	8	7	12	11	9	17	12	13	13	10
ASR -6	8	16	16	15	8	6	14	9	6	9	18	12
ASR -7	11	9	8	11	16	13	15	10	17	8	13	16
ASR -8	12	7	19	17	18	12	11	17	10	13	16	15
ASR -9	18	16	13	8	11	17	16	9	14	12	17	14
ASR -10	14	13	11	14	9	11	16	13	7	18	9	11
Mean	13.1	13.1	12.9	13.3	11.6	11.1	12.1	13	10.8	12	13.9	12.9
Std. Deviation	3.17	3.34	4.20	3.80	3.56	3.38	3.28	3.39	3.91	3.36	2.84	3.07
C.V.	0.242	0.255	0.32	0.285	0.307	0.30	0.271	0.261	0.362	0.28	0.204	0.23
Relative Importance	3	4	11	8	10	9	6	5	12	7	1	2

• Simultaneous Interpreting Test

Simultaneous interpreting test was conducted to check the assessment variations among assessors. This time 14 trainee interpreters were invited to participate in an oral translation test. These trainee interpreters involved in simultaneous interpreting session.

Their recordings were collected and delivered to six assessors to assess TQ of those trainee interpreters. The test includes two political speeches delivered in UN General Assembly as shown in the table below:

1	PM Theresa May speech to the UN General Assembly: 26 September 2018	English into Arabic	Mp3
2	President Sisi speech to the UN General Assembly: 26 September 2018	Arabic into English	Mp3

Only 12 trainee interpreters participated in this test, however; two trainee interpreters withdrew from the test before ending the recordings because they could not follow the audio within the planned time. The rest of the trainee interpreters delivered their interpretations. Interpretations were recorded with a form of mp-3 audio. Since two recordings were not clear and with a sort of distortion, only eight of them were involved in the assessment.

Six assessors were requested to assess the recordings of the eight trainee interpreters. The assessors listened, assessed the audio recordings, and returned the assessment sheets. The marks given were out of 20. The scores, the assessors and the trainee interpreters were coded and indexed into a table as shown in below in table 2.

Table (2) Marks out of (20)
ASR = Assessors TT = Trainee Translator/Interpreters

Details	TT-1	TT-2	TT-3	TT -4	TT -5	TT -6	TT -7	TT -8
ASR - 1	13	16	15	14	18	15	11	8
ASR -2	15	12	11	16	7	10	10	12
ASR -3	9	18	8	13	10	12	8	15
ASR -4	11	8	13	9	10	14	18	13
ASR -5	11	9	9	16	14	10	15	7
ASR -6	7	11	10	17	11	7	9	10
Mean	11	12.33	11	14.16	11.66	11.33	11.83	10.83
Std. Deviation	2.82	3.93	2.60	2.92	3.82	2.94	3.86	3.06
c.v.	0.257	0.318	0.237	0.206	0.328	0.259	0.326	0.282
Relative Importance	3	6	2	1	8	4	7	5

- After testing trainee translators and trainee interpreters by two successive tests: *written translation* test and *simultaneous interpreting test*, two statistical analyses of SPSS statistics were carried out to show the descriptive statistics for the assessment sheets delivered by the assessors. One sample *t-test* statistics was done for the mean value of the assessment results in order to test the hypothesis of variations among the translation assessors (Myers and A Well, 2003)

Discussion of the Results

- **Written Translation Test**

The results of the statistics represented by the coefficient of variance (CV) to the assessment sheets show clear variations in the levels of the assessors' assessments to the translation quality. This can clearly be noticed in figure (1). The higher ratio was .362 of the TT-9 and the lowest ratio was

.204 of TT-11. It is obvious that the ratio goes down for TT-9 because of the decline occurred in the ratio of the standard deviation (SD) of 2.84 with mean ratio of 13.9, and that is the highest among the rest of the trainee translators. Particularly, the numerator of the CV in question is the SD and its denominator spontaneously interprets the rise of CV of TT-9. That is because of the high convergence of the two ratios: the numerator as well as its denominator among the rest of the trainee translators. Overall, the increase of CV reflects a bigger difference and variability in the assessment process fulfilled by the assessors, while the decline in this ratio is quite the opposite one. Moreover, figure (1) perhaps clarifies this fluctuation in CV of the trainee translators' scores who responded to the test.

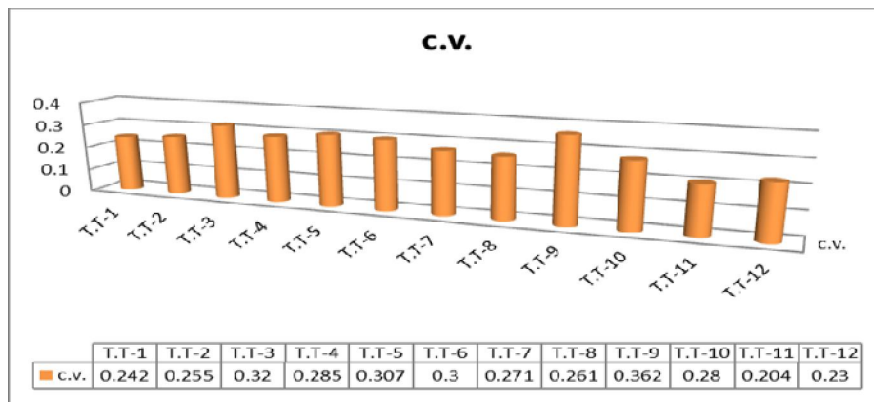


Figure (1) Statistical Description of the Written Test

Simultaneous Interpreting Test

Simultaneous interpreting test of the trainee translators/interpreters with their scores were analyzed with SPSS statistics. Thus, the task here is to summarize descriptive statistics to reveal the ratio of variations among the assessors. The statistical results showed CV via calculating the scoring variations among trainee translators. This demonstrates great variations in the assessment levels done by the assessors. The highest ratio of CV was .328 of TT-5 with highest relative importance 8. The lowest ratio of the relative importance was 1 with CV .206 of TT-4.

It is a natural result in which the rate of CV for TT-4 decreases with relative importance 14.16 after the decline in SD of 2.92 as compared to the highest mean among the other trainee interpreters. In other word, the numerator of CV is the SD and its denominator is the mean. This interprets that the CV of TT-5 is the highest among the other trainee translators since the value of the numerator and the denominator is closer to each other in their ratios. In all cases, this reflects the increase of CV ratio due to the great variation and fluctuation in the assessment levels among the assessors as shown in Figure (2). Moreover, the decline in CV refers exactly to the opposit

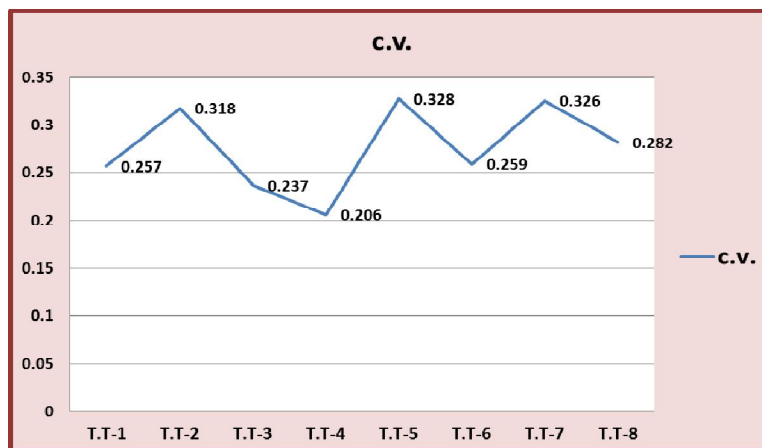


Figure (2) Statistical Description of the Simultaneous Interpreting Test

One sample T-test

Another statistical analysis was done (one sample T-Test) in order to shed light on the variations in the assessment of TQ exercised

by the translation assessors. One sample t-test is used to show the possibility of creating some observations by using certain mean (Phusewiki.org, 2011).

One sample t-test was used in this paper to compare the mean of the sample represented by the assessors in order to find out the assessment variations done in TQ. If the t-test is statistically significant, then the hypothesis of variations is definitely acceptable (Tae, 2015). For example, it is possible to use t-test to test the extent of any significant variations in the results of the assessment of translation assessors for a number of trainee translators.

In other word, there are real variations in the assessment of the translation experts. The opposite is true if the t-test is statistically not significant then it confirms that there are no variations in the assessments of TQ of trainee translators done by the translation assessors.

The review of the literature has revealed possible differences in TQ levels for several reasons. These reasons, perhaps including the difference in the results of TQA that might have contributed to factors associated with translation theories or other factors in relation to the assessor's multiple thinking styles that he/she functions during the TQA process.

Accordingly, the variations of TQ might be as a result of the interaction of several

factors with each other that can hopefully interpret the expected difference of the TQ levels during the assessment process.

In order to investigate the extent of a variation between the results of the assessment done by translation assessors to trainee translators, the hypothesis was tested, i.e. ***“There are significant variations among translation assessors in assessing the translation quality of trainee translators/interpreters”***.

One sample t-Test was done as shown in the table (3) clearly calcifies the results of testing this hypothesis at the level of written translation. Whereas, table (4) highlights the results of testing the hypothesis itself at the level of simultaneous interpreting mode. These results have proved the existence of significant variations among the results of TQA.

This could lead us to berth the ship, viz; there is an apparent existence of significant variations in the TQ based on the assessment results of the assessors who fulfilled the testing of both the written and the oral translation for the trainee translators/interpreters as shown in the figures (3) and (4).

$df = 11, (p < .01)$
Table (3)

Details	"t" Statistics	Mean Difference	95% Confidence Interval of the Difference	
			Lower	Upper
TT - 1	13.035	13.10000	10.8266	15.3734
TT - 2	12.372	13.10000	10.7048	15.4952
TT - 3	9.708	12.90000	9.8942	15.9058
TT - 4	11.062	13.30000	10.5802	16.0198
TT - 5	10.289	11.60000	9.0496	14.1504
TT - 6	10.381	11.10000	8.6811	13.5189
TT - 7	11.661	12.10000	9.7527	14.4473
TT - 8	12.093	13.00000	10.5683	15.4317
TT - 9	8.734	10.80000	8.0029	13.5971
TT - 10	11.272	12.00000	9.5917	14.4083
TT - 11	15.444	13.90000	11.8641	15.9359
TT - 12	13.282	12.90000	10.7029	15.0971

Figure(3)

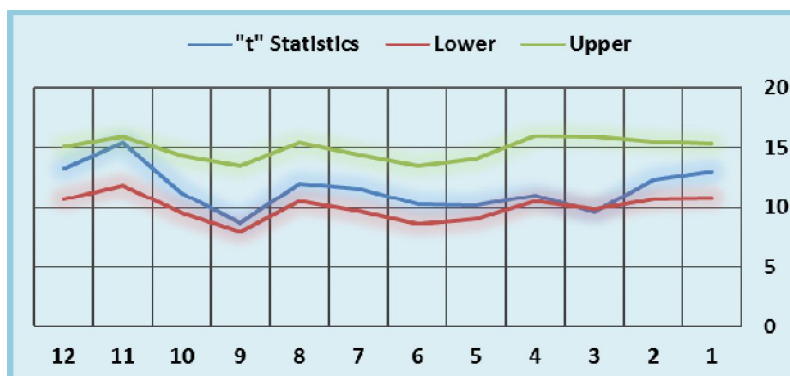


Table (4)

$df = 5, (p < .01)$

Details	"t" Statistics	Mean Difference	95% Confidence Interval of the Difference	
			Lower	Upper
TT - 1	9.526	11.0000	8.0317	13.9683
TT - 2	7.682	12.3333	8.2061	16.4605
TT - 3	10.333	11.0000	8.2634	13.7366
TT - 4	11.856	14.1667	11.0951	17.2382
TT - 5	7.862	11.6667	7.6476	15.6857
TT - 6	9.430	11.3333	8.2439	14.4228
TT - 7	7.992	11.8333	7.7734	15.8933
TT - 8	8.671	10.8333	7.6215	14.0451

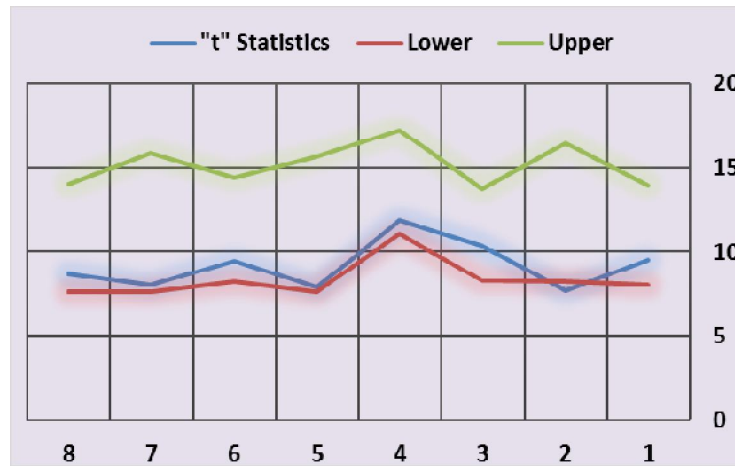


Figure (4)

Conclusion

This paper could come up with a conclusion answering the research question and proving that there are clear variations in the TQA of trainee translators/ interpreters fulfilled by some translation assessors. In addition, these variations among translation assessors might be attributed to many factors. Factors such as the way of employing translation theories in the process of TQA, the psychological effects of the assessor himself, the pressure of work, the orientations of translation institutions, or perhaps other factors may contribute in one way or another to affect the process of TQA. T-test was used to

verify the significance of the hypothesis of variation in TQA among the translation assessors for the sample of trainee translators/interpreters. The results showed that the mean value of the TQA of the translation assessors is totally variant ranges from value into another.

It is recommended to look for some factors might affect the process of TQA, such as exploring the employment of translation theories or investigating another factor such as the psychological aspects that in a way or another might affect the translation assessors when handling the assessment of TQ.



References

- Bowker, Lynne. 2000. "A Corpus-Based Approach to Evaluating Student Translations." In "Evaluation and Translation," special issue, *The Translator* 6 (2): 183–210.
- Clarke, B. (2008). Malcolm Williams. Translation Quality Assessment: An Argumentation-Centred Approach. *Cadernos de Tradução*, 1(21).
- Hang, H.T.D. (2019). An Assessment of the Vietnamese Translation of "The Adventures of Huckleberry Finn - Chapter Xx" Using House's Translation Quality Assessment Model. *VNU Journal of Foreign Studies*, 35(1).
- House, J. (1977). A Model for Assessing Translation Quality. *Meta: Journal des traducteurs*, 22(2), p.103.
- Liu, X. and Zhao, Y. (2016). The study on Translation Quality Assessment in the Translation Contests Based on Malcolm Williams' Argumentation-centered Translation Quality Assessment Model. *Proceedings of the 2016 International Conference on Humanity, Education and Social Science*. [online] Available at: <https://www.atlantispress.com/proceedings/ichess-16/25857353>.
- Myers, J.L. and A Well (2003). *Research design and statistical analysis*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Newmark, P. (2008). *A textbook of translation*. Harlow: Longman.
- Oxford English Dictionary. (2015). *Oxford English Dictionary*. [online] Available at: <https://public.oed.com/updates/new-words-list-june-2015/> [Accessed 18 Nov. 2019].
- Phusewiki.org. (2011). *Review of Statistical Analysis Plans - PHUSE Wiki*. [online] Available at: https://www.phusewiki.org/wiki/index.php?title=Review_of_Statistical_Analysis_Plans [Accessed 17 Nov. 2019].
- Reiss, K., Hans Josef Vermeer, Nord, C. and Dudenhöfer, M. (2015). *Towards a general theory of translational action: skopos theory explained*. London; New York: Routledge Taylor & Francis Group.
- Tae, K. K (2015). T Test as a Parametric Statistics. *Korean Journal of Anesthesiology*. *Korean J Anesthesiol*. 2015 Dec; 68(6): 540-546.