

## The Effectiveness of Translation Quality Assessment (TQA) Models in the Translation of Indonesian Students

Yana Shanti Manipuspika 

Study Program of English, Faculty of Cultural Studies, Universitas Brawijaya, Malang, Indonesia

✉ **Corresponding Author:** Yana Shanti Manipuspika, **E-mail:** yana\_shanti@ub.ac.id

### ARTICLE INFORMATION

**Received:** April 11, 2021

**Accepted:** May 22, 2021

**Volume:** 4

**Issue:** 5

**DOI:** 10.32996/ijllt.2021.4.5.32

### KEYWORDS

Translation, translation quality assessment, holistic approach, error analysis approach, student translation, effectiveness

### ABSTRACT

Translation Quality Assessment (TQA) has always been a subject of interest to the translation community, and with the progress of translation training and education in some parts of the world, including Indonesia, the need for translation assessment to measure students' skills also increased. This present study investigates the effectiveness of two translation assessment models in students' translation result. It aims to explore how students improve their translation after giving feedback based on two different assessment models. A mixed-method design, which is characterized by the combination of qualitative and quantitative research components, was employed. The data were collected from English Study Program students of Faculty of Cultural Studies (FCS) Universitas Brawijaya Indonesia from two translation courses at the even semester of the 2020/2021 academic year. The students were asked to translate two different types of texts, namely short story and news article. Data analysis involved providing a description of the results and performing a statistical test using SPSS, t-test in particular. The findings revealed that the students improved greatly in their translation after assessed with two different models. This can be seen from the mean scores of their translation. Waddington Method C as the holistic approach and ATA Framework as the analytical approach gave different results for different texts. Both models were similarly effective for literary text, while for journalistic text, ATA Framework was more effective than Waddington's. The recommendation is made for future researchers to combine two types of assessment models and test them to see the effectiveness. In addition, focusing on different types of source texts, such as manual text, or legal text, or academic text, or others to be translated and assessed, may also be beneficial.

### 1. Introduction

Translation quality assessment (TQA) has always been a subject of interest to the translation community. With the progress of translation training and education in some parts of the world, the need for translation assessment to measure students' skills also increased. In some Indonesian universities, translation course is offered as one of the courses in Department of Languages, one of them is Universitas Brawijaya (UB), Malang, Indonesia. In the English Study Program of UB, translation courses are divided into two, offered in two adjacent semesters; one is Introduction to Translation (compulsory course), and the other is Translation and Interpreting, an elective course.

The demand for translation nowadays is still growing, and the graduates of the English Study Program are expected to meet the standard competencies, one of which is translating competence. Due to the formulation of competencies and attainment levels for nearly every activity, it is necessary to know when a translation meets the standards that have been set and when someone can "officially" call him or herself a translator. Besides, translation quality assessment also plays a crucial role in increasing the profession's public appreciation and legal protection (Thelen, 2008).

The growth in digital content also displays broader requirements for Translation Quality Assessment (TQA), including text types, appropriate methods for the domain, workflow, as well as end-users (Moorkens et al., 2018). However, there is still a lot of misunderstanding and ignorance over the term translation quality assessment. The term does not have a clear meaning in the sense of what should be assessed and what can be assessed. It is because the process of translation is quite subjective.

The translation is characterized by some variables: knowledge/ know-how, translation skills, artistic skills ("playing with language"), and personal taste. Newmark (1988) stated that translation is a science, a skill, an art, and a matter of taste. It is a science that requires knowledge and fact verification and the language describing them. The translation is a skill, which needs appropriate language and acceptable usage. It is also an art, which differentiates good from undistinguished writing and is creative, the intuitive, occasionally the inspired, level of the translator. The translation is also a matter of taste, where the argument breaks, preferences are expressed, and various rewarding translations reflect individual differences. In reality, not all these variables can be assessed objectively, in line with the points discussed and the criteria mentioned. Thelen (2008) acknowledges that different criteria should at least assess translation results. It is referred to Newmark (1988, pp. 189-192), stating that the fourth variable (taste) can only be measured subjectively.

The relevance of translation quality assessment (TQA) is now getting stronger. Professional translators, their clients, researchers, and trainee translators need TQA for different reasons (Williams, 2009). Among the many translation quality assessment (TQA) models, most research conducted in translation quality assessment was theoretical and descriptive (Waddington, 2001). Darbelnet (1977) and Newmark (1991) are two scholars, among others, who set the criteria for a good translation. Meanwhile, House (1996), Nord (1993), and Gouadec (1981) tried to define the nature of translation errors. House (1996, as cited in Shahraki & Karimnia, 2011, p.5219) tried to build a bridge between quality assessment and text-linguistic analysis. House (2006) describes that research on texts as units larger than sentences has influenced translation studies; however, the notion of context, its relation to the text, and its role in translation has received much less attention.

This present study utilizes two translation assessment models on students' translation results. One is an error-based translation evaluation system as the procedure for quantifying the quality of translation, and the other one is using a holistic approach. Mateo (2014) stated that the product-centred methods in assessing translation results are divided into two branches. One of the trends examines the linguistic features of translated texts at the sentence level (error-based), whereas the other trend highlights macrostructure relations of the text as a unit (holistic). The ATA framework is a model used by the American Translators' Association. It is one of the models that applies the error analysis approach because it provides a detailed explanation of any errors.

One of the scholars proposing a holistic approach in TQA was Waddington (2001). He states that generally, there are three methods instructors use to evaluate student translations: error analysis, holistic approach and a combination of error analysis and holistic judgment. He believes that the sum of errors may not directly reflect the quality of translation. Teachers or instructors still need to employ the traditional and subjective criteria to assess student translations because the supposedly objective criteria proposed for evaluation are not manageable in educational contexts. Waddington (2001) proposes a scale for holistic method, which will be employed for this study to see how effective the method is to assess the improvement students make in their translation.

As stated earlier, many scholars focusing on translation have proposed criteria and/ or models for translation assessment. Albeit at various models, student translation is considered an under-researched area in evaluation. According to Medadian & Mahabadi (2015), almost none of TQA are tailor-made for a manageable summative evaluation of student translation. Therefore, most translation teachers still draw on holistic and traditional methods of translation evaluation in their exams. A holistic method of assessment is generally seen to be the best way to train translators. It is due to the fact that stressing positive grading is an appropriate way of improving students' translation quality. This means that instructors should reward for good performance rather than punishing for poor performance.

This present study explores how students of translation class can improve their translation after given feedback using two different translation assessment models. In addition, this study also tries to compare which translation quality assessment models are more effective for assessing students' translation results. Novice translators are selected since the research on TQA using students as participants are still limited. This study is worth conducting for it scrutinizes the merits of an evaluation of student translation. It focuses on the product-based summative evaluation of student translations in the undergraduate program of English at Universitas Brawijaya (UB).

This study attempts to answer the following research problems: (1) Can students of translation class improve their translation with either of the assessment models? and (2) How effective are the two TQA models in assessing students' Indonesian-English translation results? This study explores how students improve their translation results after giving feedback based on different assessment models. It also aims to reveal how effective the two models are in assessing student translation results. It is worth identifying as the information collected will be used to consider using a particular assessment scale in the translation classes at the English Department.

It is necessary to highlight that this article focuses on the assessment of student performance in translation practice course and not on assessment in general, which assess translators' competence. Therefore, assessment here means grading translation assignments instead of assessing the whole translation program.

## **2. Literature Review**

### **2.1 Translation Quality**

#### 2.1 Translation Quality

The quality of translation, fundamentally, should meet the defined criteria. It applies to those progressing in the translation industry and the students of translation. Samir & Yazdi (2020) reveal that translation quality (TQ) is a problematic concept, and defining it will not be easy. There are two definitions of TQ proposed by Koby et al. (2014), namely broad and narrow ones. The broad view of TQ defines, "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" (Koby et al., 2014, p. 416). Conversely, the narrow view states, "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." (Koby et al., 2014, p. 417).

Generally, the broad view is considered sufficient. Akbari & Shahnazari (2015) found out that the quality of translation is determined by the translator's knowledge regarding source language (SL) and target language (TL), the intention of the reader, and context (p. 445). In short, it can be stated that when the clarity, use of common language, and similarity of content and meaning are fulfilled, the translation can be called fulfilling TQ.

### **2.2 Translation Quality Assessment (TQA)**

In Translation Studies, the quality of translation is required to assess. Translation quality assessment (TQA), which is how translation quality is graded, becomes a central issue in a product-oriented translation and gets more attention from both translation scholars and experts. Several devices to assess translation quality have currently emerged, among which TQA models are described in this study. Kamalizad & Khaksar (2018) reveal that every TQA model introduces new ideas and ways to assess the translation quality integratively, discretely or combining both with respect to the theoretical contexts.

Meanwhile, Sofyan & Tarigan (2019) ascertain that TQA model used by different raters in assessing similar translation resulted in different quality. While actually, the results of using the same TQA model in assessing the same target text (TT) should show a similar level of quality of the TT although assessed by different raters. The subjectivity might cause different results in applying the TQA model or by insufficient quality criteria of the model.

Research in the field of translation quality assessment has been mainly theoretical and descriptive. Waddington (2001, pp. 16-17) determines that the research has concentrated largely on the following themes: establishing the criteria for a "good translation", 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, basing the quality assessment on text linguistic analysis, establishing various textual levels on a hierarchical basis and linking the importance of mistakes to these levels, and attempts to elaborate scales to describe different levels of translation competence.

As one of the leading scholars in the field of TQA, Waddington (2001) proposes 4 models, namely Method A, Method B, Method C, and Method D. Methods A and B are error-based, Method C is a holistic method of assessment, and Method D is a combination. His modelling is based on discussing two kinds of methods typically used at European universities, those based on error analysis and those based on a holistic approach.

#### **2.2.1 Waddington's TQA Models**

##### **2.2.1.1 Method A**

Errors are divided into three categories:

- a. Inappropriate renderings that affect the source text's understanding. These are divided into eight categories: *contresens*, *faux sens*, *nonsens*, addition, omission, unresolved extralinguistic references, loss of meaning, and inappropriate linguistic variation (e.g., register, style, and dialect).
- b. Inappropriate renderings which affect target language (TL) expression. These are divided into five categories: spelling, grammar, lexical items, text, and style.
- c. Inadequate renderings that affect the transmission of either the source text's main function or secondary functions.

A distinction is made between serious errors (-2 points) and minor errors (-1 point) in each of the categories. A fourth category describes the plus points to be awarded for good (+1 point) or exceptionally good solutions (+2 points) to translation problems. In the case of the translation exam where this method was used, the sum of the negative points was subtracted from a total of 110 and then divided by 11 to reach a mark from 0 to 10 (which is the normal Spanish system) (Waddington, 2001, p. 3). For instance, if a student gets a total of -66 points, his result would be calculated as follows:  $(110-66=44)/11=4$  (which fails to pass; the lowest passing grade is 5).

#### **2.2.1.2 Method B**

Method B is also based on error analysis and was designed to consider the negative effect of errors on the overall quality of the translations. The corrector 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: if it does not, it is a language error (and is penalized with -1 point); if it does, it is a translation error (and is penalized with -2 points).

The mark for each translation is calculated in the same way as for Method A. First, the examiner fixes a total number of positive points (in the case of method B, this was 85), subtracting the total number of negative points from this figure, and finally dividing the result by 8.5. For instance, if a student is given 30 minus points, his total mark would be 6.5 (pass):  $85-30 = 55/8.5 = 6.5$ .

#### **2.2.1.3 Method C**

This method is a holistic method of assessment. In this method, the translation competence was considered as a whole. The examiner should consider the three aspects of the translator's performance. Those are the accuracy of transfer of the source language (SL) content to the target language (TL), the quality of expressions provided in the TL, and the degree of task completion. Waddington (2001) pictures this to aid raters or teachers to judge their students' translation more consistently by providing more complete and distinguished descriptors.

In order to achieve acceptable levels of reliability, Waddington (2001) designs five levels of performance in this method. Then, he determined two possible scores for each level. In this case, if a translation fully fulfils 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 a lower score (p. 315). In addition, if a student fulfils a description at a certain level, the rater should choose between the lowest mark at that level (for example, 7 at level 4) and the highest mark at the lower level (6 at level 3). In this particular method, raters are free to use half points (5.5, 6.5), as it would then prove easier to detect possible differences by their applications of this method.

#### **2.2.1.4 Method D**

Method D provides a combination of error analysis Method B and holistic Method C in an appropriation of 70/30. It means Method B accounts for 70% of the total result and method C for the remaining 30% (p. 315).

For the purpose of this study, Waddington's (2001) Method C will be employed as the representation of the holistic assessment model and the ATA framework as the error-based model.

### **2.2.2 ATA Framework**

ATA Framework is a TQA model used by the American Translators' Association. This model applies an error analysis approach as it provides a detailed explanation of many types of errors. The ATA Framework is an assessment model to evaluate the translation results of the participants taking the test held by ATA to be certified translators. It is complex and consists of a flexible instrument containing a detailed, metric-driven error checklist (Dewi & Hidayat, 2020). It consists of 3 components: (1) a weighted matrix of error checking, (2) a chart listing the error names (labels) and descriptions of the individual errors, and (3) a flowchart guiding weighting the errors. Doyle (2003) stated that this framework provides a ready-made, standardized, time-tested, and professionally recognized model for conducting theory-based, systematic, coherent, and consistent evaluations of student translations. This framework was initially designed for certification; however, it has also been applied to evaluate translation participants or students (Dewi & Hidayat, 2020). ATA Framework can be adapted and adjusted from a product-

oriented scale into a more process-oriented scale (Koby & Baer, 2005). This model has two types of errors: (1) translation/strategic/transfer errors and (2) mechanical errors. The ATA error marking scale categories reflect the theory of Vinay and Dalbernet (1958/1995). They first came up with a list of translation errors, such as addition, omission, and mistranslation.

### **3. Methodology**

#### **3.1 Research Design**

This study applied a mixed-method approach in which the researcher collects and analyzes both quantitative and qualitative data within the same study. The quantitative approach was employed for identifying the improvement of student translation and the effectiveness of the two TQA models. T-test of SPSS was run for obtaining quantitative results. Meanwhile, the qualitative approach was used for describing the findings.

#### **3.2 Data Source**

This study was respondent-oriented, yet the focus was not on improving the respondents' skills but on discovering the effectiveness of the two assessment models. So, it is conventional or traditional research where the participants are the objects of the study (Dewi & Hidayat, 2020). It is necessary to emphasize that this study is not action research (AR) as it is not participative and cyclic. It had two phases of data collection, but it only involved gathering data and analyzing it without any prior observation, initial plan of intervention, and planning of a new intervention (Cravo & Neves, 2007). Therefore, it is not cyclic.

The data were collected from English Department students of Faculty of Cultural Studies, Universitas Brawijaya (FCS-UB) from two classes taking translation course at the even semester of 2020/2021 academic year. Novice translators were selected as this research tried to reveal the effectiveness of two different rubrics in student translation. The rater was the researcher herself as she is a lecturer of translation and a professional translator. The language pair used for this study was English – Indonesian because the students are in their third year of university, and they are in the upper-intermediate level since they have passed five semesters of English subjects comprising reading, listening, and listening speaking, writing, and grammar courses. So, their English mastery is sufficient, and they are all native Indonesians. The participants were those taking the Translation and Interpreting course in the even semester who are already equipped with sufficient knowledge on translation since they have passed the Introduction to Translation course in the earlier semester.

#### **3.3 Data Collection**

The students translated two different texts from English into Indonesian. One of the source texts was a short story. The selection of a literary text was based on the consideration that machine translation would not be any help. Since the data collection was conducted online due to the pandemic, the researcher selected a text which could not be translated by machine translation. The short story is entitled 'Later' by Michael Foster. The other text used for students to translate was a journalistic text taken from BBC news entitled 'Covid vaccines: Why some Americans are choosy about their jab' written by Cache McClay accessible via <https://www.bbc.com/news/world-us-canada-56410179>. This topic is considered relevant to the situation at the time this study was conducted.

For each text type, the first data was obtained by assessing students' translation using the two different assessment models (Waddington's Method C and ATA) and each result was given feedback then the errors were identified and counted. After students revised their work based on the feedback, the second data was taken by after students revised their work.

#### **3.4 Data Analysis**

After the rater assessed the students' translation, the errors found in the first and the revision results were compared to see the improvement. The number of errors in the first draft of student translation was counted, then it was compared to the errors in the revised version. In analyzing data, a statistical test using SPSS (Statistical Package for the Social Sciences) was performed to find out whether or not the difference in the effectiveness between the two assessment models is significant. The results were then described and a conclusion was drawn.

### **4. Results and Discussion**

#### **4.1 Results**

##### **4.1.1 Students' Translation Improvement**

The results of this study were obtained by analyzing students' translation both manually and statistically using SPSS. The SPSS assists in indicating the more effective assessment models between the two.

In the following analysis, text 1 refers to a literary text (the short story), and text 2 is the journalistic text. The improvement was observed from the difference of errors in the revised version of the translation. The errors include both translation and language errors. This study follows Dewi and Hidayat's (2020) in deciding the errors in student translation. The number of translation errors

(TE) was subtracted from the number of revision errors (RE), and this generated an error difference (ED). The formula is as follows:

$$TE - RE = ED$$

With that formula, the first text (literary text type) data in the form of short story is displayed in Table 4.1.

Table 4.1 Improvement Results of Text 1

Participant number	Waddington's Model C				Participant number	ATA Framework			
	TE	RE	ED	ED (%)		TE	RE	ED	ED (%)
001	6	0	6	100	021	20	8	12	60
002	12	6	6	50	022	11	6	5	45.5
003	6	3	3	50	023	6	3	3	50
004	5	2	3	60	024	5	2	3	60
005	5	1	4	80	025	5	3	2	40
006	9	3	6	66.7	026	10	5	5	50
007	11	6	5	45.5	027	3	0	3	100
008	6	2	4	66.7	028	4	0	4	100
009	11	0	11	100	029	8	3	5	62.5
010	7	3	4	57.1	030	6	3	3	50
011	7	2	5	71.4	031	5	1	4	80
012	6	2	4	66.7	032	12	5	7	58.3
013	9	0	9	100	033	4	1	3	75
014	3	0	3	100	034	7	2	5	71.4
015	15	4	11	73.3	035	11	5	6	54.5
016	11	4	7	63.6	036	10	5	5	50
017	6	1	5	83.3	037	5	0	5	100
018	9	2	7	77.8	038	5	1	4	80
019	4	2	2	50	039	7	3	4	57.1
020	9	3	6	66.7	040	5	2	3	60
<b>Average</b>	<b>7.9</b>	<b>2.3</b>	<b>5.6</b>	<b>71.4</b>	<b>Average</b>	<b>7.5</b>	<b>2.9</b>	<b>4.6</b>	<b>65.2</b>

Table 4.1 displays the data of improvement results from Literary text. Among 40 participants, 50% were given feedback by employing Waddington's Model, and the other 50% using ATA Framework.

The table shows that the improvement results of literary text translation with Waddington's Model ranged from 45.5% - 100% decrease of errors. Below 50% improvement was found only in one data, and the other 19 data got 50% improvement or higher. On average, the improvement was 71.4%. Meanwhile, the average improvement of translation assessed by the ATA Framework indicated 65.2%. Regarding the range of errors, those given feedback using ATA Framework show a similar result. It ranged from 45.5% - 100% decrease of errors. Only 1 data showed <50% improvement. This finding shows a great improvement of literary text translation results in the revisions using both assessment models, considering that majority (>50% data showed >50% improvement).

The same formula was used to count for the difference of errors in another text type, journalistic one. The results can be seen in Table 4.2.

Table 4.2 Improvement Results of Text 2

Participant number	Waddington's Model C				Participant number	ATA Framework			
	TE	RE	ED	ED (%)		TE	RE	ED	ED (%)
021	10	5	5	50	001	10	2	8	80
022	9	3	6	66.7	002	14	6	8	57.1
023	13	6	7	53.8	003	18	9	9	50
024	10	5	5	50	004	20	7	13	65
025	15	5	10	66.7	005	30	13	17	56.7
026	10	7	3	30	006	25	8	17	68
027	10	3	7	70	007	15	4	11	73.3
028	8	5	3	37.5	008	15	5	10	66.7
029	15	4	11	73.3	009	15	6	9	60
030	15	7	8	53.3	010	13	8	5	38.5
031	5	0	5	100	011	21	9	12	57.1
032	0	0	0	32.1	012	8	0	8	100
033	9	4	5	55.6	013	5	0	5	100
034	15	8	7	46.7	014	10	4	6	60
035	25	13	12	48	015	21	10	11	52.4
036	15	7	8	53.3	016	10	4	6	60
037	10	3	7	70	017	16	7	9	56.3
038	10	5	5	50	018	10	4	6	60
039	9	4	5	55.6	019	25	15	10	40
040	13	5	8	61.5	020	6	2	4	66.7
<b>Average</b>	<b>12.7</b>	<b>5.9</b>	<b>6.8</b>	<b>56.2</b>	<b>Average</b>	<b>15.4</b>	<b>6.2</b>	<b>9.2</b>	<b>63.4</b>

Table 4.2 shows that there were 20 data assessed for journalistic text with Waddington's rubric and 20 data with the ATA Framework. The improvement results of journalistic text translation assessed with Waddington's rubric ranged from 32.1% – 100%. There were 5 data showing improvement of less than 50%, while there were 15 data showing improvement of 50% and above. Likewise, the improvement of journalistic text translation assessed with ATA Framework is also substantial, as it ranged from 38.5% - 100%. Only 2 data indicated less than 50% improvement, whereas the remaining 18 data showed above 50% improvement.

We can see from the findings that it is obvious that most participants have improved their translation greatly, whether they received the Waddington's scale or ATA Framework as the assessment model. For the results assessed with Waddington's scale, the average of Text 1 was 71.4%, whereas the average of Text 2 was 56.2%. In addition, for the results assessed with ATA Framework, the average of Text 1 and Text 2 was 65.2% and 63.4%, respectively.

If seen from the average, the results assessed with Waddington's Model were higher in Text 1, and ATA Framework functioned better in Text 2. Different types of text resulted differently because the level of difficulties might be different for the participants. For revealing the difference in the effectiveness of the scales, it is important to calculate using a t-test with SPSS.

#### 4.1.2 The Effectiveness of the Assessment Models

The calculation using t-test is important to reveal whether there is a significant difference in the effectiveness between the results assessed with Waddington's Model (as a holistic approach) and those with the ATA Framework (error-analysis based). Heiman (2001, p. 393) states, "T-test is for testing a single-sample mean when (a) there is one random sample of interval or ratio data, (b) the raw score population is normally distributed, and (c) the standard deviation of the raw score population is estimated by computing  $s_x$  [standard deviation] from the sample data."

The t-test used in this study was the independent sample t-test as there were two independent variables: Waddington's Model and ATA Framework. The improvement results can be seen in Table 4.3 as follows.

Table 4.3 The Independent Sample t-test for Literary and Journalistic Texts

Group Statistics					
	Group	N	Mean	Std. Deviation	Std. Error Mean
literary	A	20	4.5500	2.13923	.47835
	W	20	5.5500	2.50210	.55949
journalistic	A	20	9.2000	3.62157	.80981
	W	20	6.3500	2.81490	.62943

\*A : ATA Framework ; W : Waddington's Model

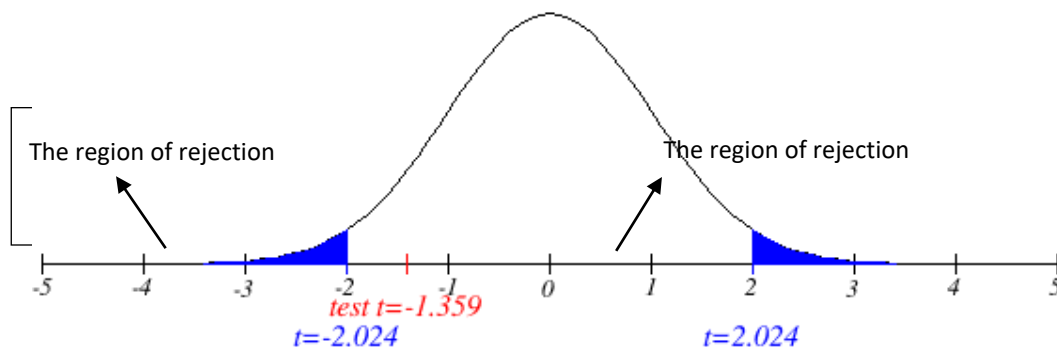
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
literary	Equal variances assumed	1.012	.321	-1.359	38	.182	-1.00000	.73610	-2.49016	.49016
	Equal variances not assumed			-1.359	37.104	.183	-1.00000	.73610	-2.49134	.49134
journalistic	Equal variances assumed	.842	.365	2.779	38	.008	2.85000	1.02566	.77366	4.92634
	Equal variances not assumed			2.779	35.819	.009	2.85000	1.02566	.76950	4.93050

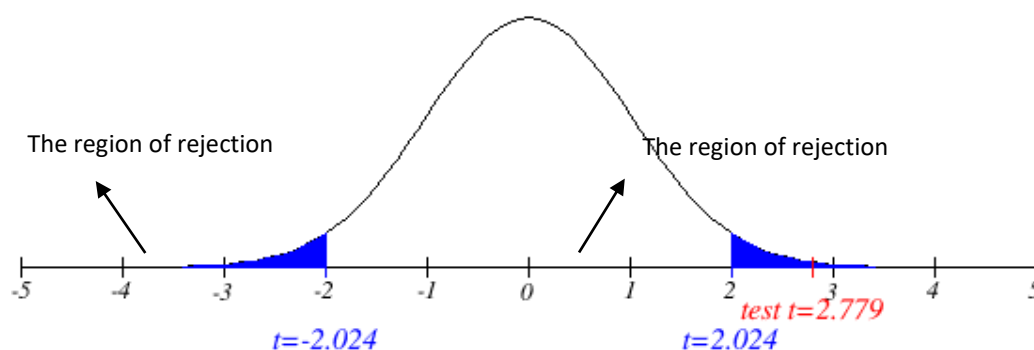
Table 4.3 shows that the total number of respondents (**N**) was 40; yet, what determines the appropriate distribution of a study is not the N, but the **df**, instead. Degrees of freedom (df) has a formula of N-2 (in the results, it was 38 for both models). The results of t test are related to whether or not the hypothesis is accepted. In this study, the hypothesis was that there is a significant difference between the translation assessed with Waddington's Model and that with ATA Framework. The translation assessed with ATA Framework shows greater improvement, and in turn, this model is more effective. The null hypothesis ( $H_0$ ) was no significant difference in effectiveness between Waddington's Model and ATA Framework as the assessment models in the English into Indonesian translation improvement. According to Heiman (2001, p.363), the null hypothesis is the statistical hypothesis that describes the population mean being represented if the predicted relationship does not exist.

The following figures show the position of t.

Figure 4.1 Sampling Distribution of Literary Text





**Figure 4.2 Sampling Distribution of Journalistic Text**

The blue areas are the region of rejection. Region of rejection is the part of a sampling distribution containing values that are so unlikely to occur that we 'reject' that they represent the underlying raw score population.  $t_{crit} = 2.024$  is the critical (significant) value of  $t$ , which indicates the boundary of the significant value of a sample mean.

Figure 4.1 shows that for Text 1 (literary text), the  $t$ -value  $-1.359$  ( $t=-1.359$ ,  $p>0.05$ ) does not lie in the region of rejection, indicating that the null hypothesis cannot be rejected (or the hypothesis is rejected). In other words, the improvement in the translation results of literary text between the ones assessed with Waddington's Scale and those with ATA Framework do not indicate any significant difference. It can be inferred that for translating literary text, a short story, in particular, the respondents can improve their work whether they are assessed using Waddington's or ATA Framework.

Concerning the improvement in translating journalistic text (Text 2), the  $t$ -value is  $2.779$  ( $t=2.779$ ,  $p<0.05$ ), which lies in the region of rejection. It means that there is a significant difference in the effectiveness of rubrics in the translation improvement of Text 2. For translating journalistic text, ATA Framework is more effective as an assessment model because the improvement in the translation results is greater than the other. With a higher mean score (mean=9.2000) than Waddington's Model (mean=6.3500), the ATA Framework is better to assess journalistic text translation.

#### 4.2 Discussion

This study investigates how novice translators (translation students, in this study) can improve their work after receiving feedback taken from two different assessment models. In addition, this study also highlights which assessment model is more effective to be used in different types of texts.

The findings have revealed that the improvement of translation in both texts, literary and journalistic, using either Waddington's Model or ATA Framework as the assessment models, was different. The  $t$ -test revealed that for translating literary text (short story), both assessment models were effective to use. However, for translating journalistic text (news article), it is more effective to use ATA Framework to assess the translation because the improvement would be greater.

Despite the fact that the quality of translation is not only about the work being free of errors, novice translators still need to consider the quality assurance of their work. Proofreading and editing processes of the work before the submission is required. For short story translation, the students involving in this study can improve their work when assessed with a holistic method. One of the advantages of holistic assessments, for instructors in particular, is that they can be utilized to evaluate various aspects of student translation. Biggs & Tang (2007, as cited in Williams, 2013, p. 440) mention that a valid assessment must be of the student's total performance, but at the same time, the conceptual framework underlying assessment, at the same time should relate the whole to its parts. The holistic assessment model does not abandon the quantitative dimension of assessment; it combines the quantitative to qualitative dimension. The quantitative element is the one facilitating reporting and justification of grades.

When translating a short story, the students made improvement well either assessed using a holistic model or error-based model. This could possibly happen because translating a short story as a form of literary text requires the modification of translators to send the message of the story across. The literal translation is not applicable for translating a short story. Doing free and communicative translation will help the translator to compose a natural translation to the target language. Translation students must be aware of style, symbols, and 'atmosphere' in short stories. Once the translator has uncovered the meaning, it is

possible to find combinations and substitutions which are faithful to the atmosphere in the original text. Since creativity is necessary for translating literary text, improvement can be made well, no matter the assessment model.

It would be different for the case of journalistic text translation. It is evident from this study that students can have better improvement when assessed with an error-based model of assessment. The journalistic translation is a field requiring experience using various techniques based on the context of the subject in hand and in-depth knowledge of both the source and target languages. Translating journalistic text needs skills and the ability to control the different documentary sources of information in order to avoid misunderstandings in the translated texts. A good journalistic translation conveys sense or proper feeling of source text and accuracy of meanings to result in original-like texts. When translating this type of text, students must take into consideration the following: the play on words, idiomatic expressions and historic referrals used in the source texts. This can be the underlying reason why ATA Framework as an error-based model of assessment is proven to be more effective to use.

The results of this study are different from Turner et al.'s (2010) research results and Dewi and Hidayat's (2020) study when it deals with the effectiveness of translation assessment models. Dewi & Hidayat (2020) revealed that the holistic approach and analytical approach used in assessing academic text are equally effective. They also discovered that there is no significant difference in the effectiveness of the two models. Likewise, Turner et al. (2010) found that holistic system assessment and analytical one has similar effectiveness. This study's different finding results in the journalistic text translation since in assessing journalistic text, ATA Framework was more effective than Waddington's Model.

This study displays similar results to Amini's (2018) research, which evaluated students' translation through three TQA models. Amini's (2018) results of statistical analysis indicated that the error analysis method B was more reliable than holistic method C. This is in line with this study that the error-analysis based assessment model is more effective for improvement. However, Amini (2018) also revealed that a combination of error analysis and holistic method resulted in a better reliability rating and more accurate results than holistic and analytic methods alone. Therefore, based on Amini's (2018) study, the combined method is suggested to be a reliable method for evaluating and scoring students' translations.

The need for a more representative TQA model, as proposed by Sofyan & Tarigan (2019), was raised due to the presence of relativity and subjectivity in a translation quality assessment. Their study found that TQA should be based on a holistic method whose model should clearly distinguish quality criteria. With the nature of the texts being translated, this present study found the opposite. According to Sofyan & Tarigan (2019), their proposed model prioritizes accuracy as the aspect of quality realized through good translation and linguistic skills. Those two elements resulted in five translation quality aspects, namely accuracy, meaning equivalence, translation skill, text function, and grammar and style.

## **5. Conclusion**

This study shows that students of translation class in FCS-UB had significant improvement in their translation after assessed with two different models. Waddington's Method C as the holistic approach and ATA Framework as the analytical approach gave different results for different texts. For translating literary text, both models were equally effective to use, while for journalistic text, ATA Framework was more effective than Waddington's. The findings confirm Amini's (2018) work that the analytical approach assessment is more effective than a holistic one; yet, this applied only to journalistic text. Whereas for the literary text, the results of this study were in line with Dewi and Hidayat's (2020) and Turner et al.'s (2010) study that both holistic system assessment and analytical system are effective.

This study is limited to the population of novice translators (students), and the texts being translated were only a short story and a news article. Future researchers are suggested to combine two types of assessment models and test them to see the effectiveness. Additionally, future research can focus on different types of source texts, such as manual text, or legal text, or academic text, or others to be translated and assessed. It will enrich research in translation assessment and establish an appropriate assessment model for English to Indonesian translation.

## **References**

- [1] 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.
- [2] Amini, M. (2018). How to evaluate the TEFL students' translations: through analytic, holistic or combined method? *Language Testing in Asia*, 8(10),1-8.
- [3] Cravo, A. & Neves, J. (2007) Action Research in Translation Studies. *The Journal of Specialised Translation*, 7, 92-107.
- [4] Dewi, H.D. & Hidayat, R.S. (2020). The Effectiveness between Two Translation Assessment Models for English to Indonesian Translation of Undergraduate Students. *Journal of Language and Literature*, 20(2),270-287.

- [5] Doyle, M. S. (2003). Translation Pedagogy and Assessment: Adopting ATA's Framework for Standard Error Marking. *The ATA Chronicle*, 32(11), 21-28 & 45.
- [6] Heiman, G.W. (2001). *Understanding Research Methods and Statistics*. Boston, MA: Houghton Mifflin Company.
- [7] House, J. (1996). *Translation Quality Assessment: A Model Revisited*. Tübingen: Narr.
- [8] House, J. (2006). Text and Context in Translation. *Journal of Pragmatics*, 38(3), 338-358.
- [9] House, J. (2015). *Translation quality assessment: Past and present*. London and New York: Routledge.
- [10] Kamalizad, J. & Khaksar, M. (2018). Translation Quality Assessment of Harry Potter and the Cursed Child according to House' TQA Model. *DJ Journal of English Language and Literature*, 3(2), 11-18.
- [11] Koby, G. S., Fields, P., Hague, D. R., Lommel, A., & Melby, A. (2014). Defining translation quality. *Tradumàtica*, (12), pp. 413-420.
- [12] Koby, G. S. & Baer, B. J. (2005) From Professional Certification to the Translator Training Classroom: Adapting the ATA Error Marking Scale. *Translation Watch Quarterly*, 1, Inaugural Issue, 33-45.
- [13] Mateo, R.M. (2014). A Deeper Look into Metrics for Translation Quality Assessment (TQA): A Case Study. *Miscelánea: A Journal of English and American Studies*, 49, 73-94.
- [14] Medadian, G. & Mahabadi, D.N. (2015) A Summative Translation Quality Assessment Model for Undergraduate Student Translations: Objectivity Versus Manageability. *Studies about Languages*, 26, 40-54.
- [15] Moorkens, J., Castilho, S., Gaspari, F., Doherty, S. (2018). *Translation Quality Assessment: From Principles to Practice*. Springer International Publishing AG.
- [16] Shahraki, A. & Karimnia, A. (2011). Waddington's model of translation quality assessment: a critical inquiry. *Elixir Linguistics and Translation*, 40, pp. 5219-5224.
- [17] Sofyan, R. & Tarigan, B. (2019). Developing a Holistic Model of Translation Quality Assessment. *Advances in Social Science, Education and Humanities Research*, 254, 266-271.
- [18] Thelen, M. (2008). Translation Quality Assessment or Quality Management and Quality Control of Translation? *Translation and Meaning*, 8, 411-424.
- [19] Turner, B., Lai, M., & Huang, N. (2010). Error deduction and descriptors – A comparison of two methods of translation test assessment. *Translation & Interpreting*, 2(1), 11-23.
- [20] Waddington, C. (2001) Different Methods of Evaluating Student Translations: The Question of Validity. *Meta*, 46 (2), 311-325.
- [21] Williams, M. (2009). Translation Quality Assessment. *Mutatis Mutandis*, 2(1), 3 – 23.
- [22] Williams, M. (2013). A holistic-componential model for assessing translation student performance and competency. *Mutatis Mutandis*, 6 (2), pp. 419-443.