The Effectiveness between Two Translation Assessment Models for English to Indonesian Translation of Undergraduate Students

Haru Deliana Dewi & Rahayu Surtiati Hidayat
harudd.dewi7@gmail.com & rahayu.surtiati@gmail.com
Linguistics Department, Faculty of Humanities, Universitas Indonesia, INDONESIA

Abstract

Research on translation assessment on English to Indonesian translation results using two dissimilar rubrics and a quantitative approach is rarely conducted by Indonesian scholars. This present study investigated the effectiveness between two assessment models, which are very different, one using a holistic approach (the LBI Bandscale) and the other using the error analysis approach (the ATA Framework). The research has been conducted on several language pairs, including the Indonesian-English translation, but it has never been done on the English-Indonesian translation. The research aims to discover whether there is a substantial improvement using both assessment models and whether one model is more effective than the other. The study was conducted in the Introduction to Translation (DDPU) classes of the English Studies Program of the Faculty of Humanities (FIB), Universitas Indonesia (UI) for undergraduate students of Semester 6. The respondents were asked to do translation in class, and then within three weeks, their works were returned with feedback based on both models. After that, they were asked to do revisions of their translation results. The outcome of the analysis shows that there is a great improvement in the translation results because of the two assessment models, but there is no significant difference in the effectiveness between those models.

Article information

Received:
28 May, 2020

Revised:
27 June, 2020

Accepted:
6 July, 2020

Keywords: *LBI bandscale; ATA framework; assessment models; effectiveness*

DOI: 10.24071/joll.v20i2.2622
Available at https://e-journal.usd.ac.id/index.php/JOLL/index

Introduction

With the progress of translation training and education in some parts of the world, the need for translation assessment to measure students' skills has increased as well. Although in Indonesia translation training and education have not flourished as much as in countries like the US, Canada, Australia, China, some European countries, and Arab countries, several universities have started to pay attention to this field of study. One of them is Universitas Indonesia (UI), which offers a translation course at the undergraduate program and a translation major at the Master's degree program of the Faculty of Humanities (FIB). Moreover, this faculty has an institution offering translation and

interpreting services and training. In addition, there is an increasing demand for having a good translation assessment model to measure translation students' skills.

An institution under FIB UI named LBI (*Lembaga Bahasa International* – International Language Institution) has developed a translation assessment model called the LBI Bandscale, which is used to evaluate the performance of the students taking translation courses or training at LBI. This model provides an assessment scale or a holistic evaluation instrument as it gives a general description consisting of positive and negative feedback on translation results. This assessment model was formulated in 2008 as a tool to evaluate final student translations, and it is still used now. Thus, this bandscale functions as a summative assessment to decide the performance level of LBI translation students and whether they pass or fail the translation class they take. This model contains four (4) levels of measurements, applying the letters A, B, C, D to reflect the accomplishment of students, where A shows the highest achievement, and D shows the lowest. In addition to the letters, there are also comments in the bandscale which describe the positive aspects of student translation results and the errors made. Those comments serve as general, not individual or tailor-made, feedback for the participants on their final translation results. The errors mentioned are also very general and not in a detailed list. The LBI Bandscale can be seen in Appendix 1.

This type of bandscale is similar to descriptors explained by Turner et al. (2010) or a holistic approach applied by Waddington (2001a, 2001b, 2003). Turner et al. state that descriptors have been pioneered and used by organizations in the language field to assess general language competence (2010, 15). They describe what language class participants can do and what they cannot; in other words, descriptors explain the positive aspects and the negative ones of students or participants' ability or achievement. Such a system in translation is called an assessment scale or a holistic evaluation instrument, as it contains positive and negative descriptions of overall competence. According to Conde, a holistic evaluation instrument is defined as an instrument which classifies "each translation into any of the pre-defined levels within a scale" (2013, 98). The focus of this assessment model is more on translator competence than on translation products.

However, that kind of model is not the only type of translation assessment model in the world, as there is another type called the ATA Framework, which is a 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, comprising three 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 providing guidance on weighting the errors. Doyle asserts that "the framework provides a ready-made, standardized, time-tested, and professionally recognized model for conducting theory-based, systematic, coherent, and consistent evaluations of student translations" (2003, 21). This framework was initially designed for certification; however, it has also been applied to evaluate translation participants or students. Koby and Baer explain that the ATA Framework can be adapted and adjusted from a product-oriented scale into a more process-oriented scale (2005, 43). This model has two types of errors: (1) translation/strategic/transfer errors and (2) mechanical errors. The categories in the ATA error marking scale reflect the theory of Vinay and Dalbernet (1958/1995), who first came up with a list of translation errors, such as addition, omission, mistranslation, etc.

This assessment model in the translation industry is known to apply a checklist of errors, an error analysis/deduction approach, or an analytic evaluation instrument. It evaluates a translation product by counting and discovering the errors it has. Conde defines an analytic evaluation instrument as an

instrument that is based on errors, and the total grade is usually described, quantified, and subtracted from the total errors (2013, 98). This model focuses more on translation products (translation results) than on translator competence. However, it can also be applied for translation pedagogy (Arango-Keeth & Koby, 2003). The ATA Framework is usually updated, albeit not annually, and the paper applied the 2017 one, which was the newest one when the research was conducted. This framework can be seen in Appendix 2.

Because of the very different nature of these two translation models, the present paper aims to discover the effectiveness of the models to help translation students improve their translation skills or reduce errors in their translation results. There are several studies that have also researched the types of translation assessment models. In his research, Waddington (2001a, 2001b, and 2003) assessed his students' translation results from Spanish to English using four assessment models: two using an error analysis approach, one applying a holistic approach, and the last one using the combination of the error analysis and holistic approaches. He discovered that the error analysis-based model is more reliable than the holistic approach, and the combination of the approaches would produce more accuracy. Conde in his dissertation (2009) reported on evaluations conducted by professional translators, translation teachers, translation students, and potential addresses. He confirms that holistic models were as effective as error analysis models for measuring translation quality (2009, 2011, 2012a, 2012b). Turner *et al*. (2010) examined two types of assessment models, error analysis and descriptors (holistic approach), used for translating and interpreting accreditation tests. Their result indicates that both types of assessment are as reliable and accountable as translation assessment models, and both have the same level of accuracy. Dewi in her dissertation (2015) has measured the effectiveness between the LBI Bandscale and the ATA Framework in the Indonesian to English translation results and discovered that there is no significant difference in the effectiveness of the two models. Dewi (2015) applied the 2011 ATA Framework in her research, while this paper uses the 2017 one.

Both translation models are usually used as summative assessment instruments, but in this present study they function more as formative assessment instruments, as the respondents were given a chance to improve their translation. Summative assessment refers to the assessment held at the end of a learning period and decides whether students pass or fail, while formative assessment is held during the process of learning where students still have a chance to improve their performance after finding out their assessment result (Qu & Zhang, 2013). This present paper is a further study to seek for both models' effectiveness in English to Indonesian translation results, and the research of such a topic in this language pair direction has never been conducted in Indonesia. We cannot just assume that the same result from the Indonesian to English translation will be yielded too from the English to Indonesian translation without any research done on it to support the assumption. Although this present research has the same steps and focus as Dewi's (2015), the number of data is more substantial than that of the previous work, so more conclusive and solid findings can be obtained. This study attempts to answer the following questions: (1) can students improve their translation with either of the assessment models? (2) which assessment model is more effective, the LBI Bandscale or the ATA Framework, in assessing the English – Indonesian translation results?

Methodology

This section will discuss the research design, the subjects of research, data and data source, procedure of collecting data, and procedure of analyzing data.

The research applied both descriptive quantitative and qualitative methods in analyzing the data obtained. The quantitative approach is applied in discovering the improvement of student translation results and the effectiveness between two assessment rubrics used, while the qualitative approach is for the description of the findings. Data were collected from 5 April 2018 to 27 April 2018

from the English Studies Program undergraduate students of semester six who took DDPU (Introduction to Translation) classes of the Faculty of Humanities (FIB), Universitas Indonesia in Depok, West Java, Indonesia.

We chose students or novice translators as subjects of research since, in the previous research Dewi (2015) also chose novice translators. The research involving professional translators will be the next step of this type of research. The number of subjects of research was 64 out of 65 students who were willing to be the participants of the research conducted in DDPU classes. There were two classes, namely Class A and Class B, where Class A was conducted on every Thursday in the morning from 8 AM to 11 AM, and Class B was from 11 AM to 2 PM on the same day for the Second Semester around four months. Class A consisted of 33 students, all of whom participated in the research, and Class B had 32 students, 31 of whom were willing to participate in the research. In class, they studied the basic translation theory and analyzed some translation results with the translation methods and procedures taught. These students' English skills are superior and in the upper-intermediate level as they already acquired six semesters of English subjects consisting of grammar, listening and speaking, reading, and writing. All of them are native Indonesians.

Class A was taught by Mr. Andika Wijaya, M.TransIntrep., who earned his Master's degree at RMIT (Royal Melbourne Institute of Technology) University, Australia, and a professional translator with a NAATI (National Accreditation Authority for Translators and Interpreters) certification. He was the rater of the translation results of Class B. Meanwhile, Class B was taught by Haru Deliana Dewi, Ph.D., who completed her doctoral program at Kent State University in Kent, Ohio, USA, majoring in Translation Studies and a professional translator too. She was the rater for Class A translation results. The switch was conducted to make sure that the teacher of the class is not the rater of his/her class to avoid biases and power-relation. Every once or twice a week the raters met to discuss the researched translation results to have the same perception on which is considered as translation and language errors and which is not; therefore, the correlation coefficient test was not conducted for this research as both raters already agreed upon the same criteria of errors found.

The participants read, understood, and signed a consent form before they joined the research. They were very eager to participate in this study by doing the translation and the revision seriously. Only one student from Class B could not participate due to his condition. The consent form can be seen in Appendix 3. After they signed the consent form, they were provided with the instructions to work on the translation and the source text, and three weeks later, they were given the instructions to work on the revision along with their assessed translation results, the LBI Bandscale, and the ATA Framework. All the instructions and the source text can be seen in Appendix 4.

The source text was taken from TOEFL reading practice which is accessible to the public via *https://www.examenglish.com/TOEFL/TOEFL_reading1.htm*. We selected this type of text because we believe it is more familiar for students as it is an academic text, and the topic is about language, which is what students of FIB UI major in. The level of this reading is appropriate with semester six undergraduate students whose English skills are in B2 or upper intermediate level. The title of the text is *The Creator of Grammar* about the importance of grammar in language and about who creates grammar. It consists of five paragraphs, 702 words, but the participants were asked to translate only two paragraphs, paragraphs three (3) and four (4). The reasons why we selected paragraphs 3 and 4 are; first, we believe paragraph 1 should be part of the background to make the students understand the reading before translating, so it should not be part of the assignment; second, only paragraphs 3 and 4 have almost similar number of words and are in a row of each other, while paragraph 2 and 5 are very short each. Paragraph 3 of this essay, hereinafter referred to as Paragraph One, consists of 176 words, and paragraph 4, hereinafter referred to as Paragraph Two, consists of 193 words.

Prior to explaining the procedure of this research, it is important to emphasize that this study is not action research (AR) as proposed by Cravo and Neves (2007) since it is not participative and cyclic. It is a participant-oriented study, but the focus is not on improving the skills of the participants but on discovering the effectiveness between two assessment models, so it is conventional or traditional research where the participants are objects of the study. It has two phases of taking the data, but it only involves data collection and analysis without any prior observation, initial plan of intervention, and planning of a new intervention (Cravo and Neves, 2007, 94); thus, it is not cyclic. Dick (1993) explains that in a traditional piece of research the data are collected first and then analysis is carried out, while in AR the data of the study can be improved first significantly by combining data collection, interpretation, library research, and reporting. Obviously, based on those scholars' opinions, this present study is definitely not AR.

The research was announced to the students of DDPU classes from the beginning of the semester in February 2018. The researchers conducted the study after the mid-term test was done so that the students were already equipped with sufficient knowledge on translation. On 5 April 2018, the study started in class by asking the students to read and understand the consent form before they were willing to sign it. After they signed the consent form, they were given the instructions to do the translation of two paragraphs and the source text in one file. They were given one hour to complete the translation and they were able to look at any resources to assist them with the translation, except asking questions to their peers and teacher. When they finished, they sent the results to the teacher via email. The Class A teacher sent the translation results to the Class B teacher to be assessed, and the Class B teacher sent them to the Class A teacher. This is how we obtained the first data.

The second data were obtained after students revised their translation based on the feedback provided using the two assessment models, the ATA Framework and the LBI Bandscale. The participants did the revision also in class on 27 April 2018 for one hour with the permission to look at any resources, except asking questions to their peers and teacher, and when they finished, the revision results were sent by email to the class teacher. The Class A teacher sent them to the Class B teacher, and vice versa. The participants were given a small token of appreciation in a form of a keychain. The raters then completed the data collection and did the analysis.

The raters had three weeks to assess the translation results using the LBI Bandscale for 50% done on Paragraph One and 50% on Paragraph Two, and using the ATA Framework for 50% done on Paragraph One and 50% on Paragraph Two. The participants' names were coded, so the raters could not link the results to the students, or they would be anonymous. Translation errors based on the list of errors in the ATA Framework were discovered and counted in participants' translation results, and each result was given feedback. The feedback provided depended on which rubric was applied in the paragraph translated. If the paragraph was assessed using the ATA Framework, then the feedback was given based on the rubric. If it was assessed using the LBI Bandscale, then the feedback was also based on that rubric.

When the second data (the revision results) were received by the raters, they were analyzed similarly like the first data (the translation results) where the errors were discovered and tallied. Then the number of errors in the revision results (the second data) was compared with the number of errors in the translation results (the first data). From there, we could find whether participants improved or not in their translation. The improvement with the ATA Framework was compared with the improvement with the LBI Bandscale. SPSS (Statistical Package for the Social Sciences) was applied to discover whether the difference in the effectiveness between the two assessment models was significant or not.

Results and Discussion

The research results were analyzed manually and with SPSS. The manual analysis is to show whether there is an improvement from the revisions using the LBI Bandscale and the ATA Framework. The improvement refers

to the decrease of the errors discovered in the revisions compared to the number of errors in the translation results. The SPSS helps to show which assessment model is more effective for the participants to improve.

Improvement with the Assessment Models

The improvement here is observed from the decrease of errors found in the revisions of the translation results. The errors include translation and language errors, and they are counted as one for a single error with no different weighted scale. If the same error is repeated five times, for example, found in a translation result, then it is counted as five errors. The number of translation errors (TE) is subtracted from the number of revision errors (RE), and this generates the error difference (ED). The following is the formula:

TE – RE = ED

With the formula, the following table shows the data of Paragraph One from DDPU classes.

Table 1. Paragraph One Improvement Results

Respondent	The LBI Bandscale				Respondent	The ATA Framework			
	TE	RE	ED	ED%		TE	RE	ED	ED%
A12018	8	4	4	0%	A22018	11	11	0	0%
A32018	15	4	11	73.3%	A42018	7	5	2	28.6%
A52018	20	10	10	50%	A62018	15	3	12	80%
A72018	11	9	2	18.2%	A82018	14	6	8	57.1%
A92018	7	2	5	71.4%	A102018	9	6	3	33.3%
A112018	7	2	5	71.4%	A122018	17	4	13	76.5%
A132018	21	14	7	33.3%	A142018	10	3	7	70%
A152018	16	6	10	62.5%	A162018	11	4	7	63.6%
A172018	13	3	10	76.9%	A182018	16	6	10	62.5%
A192018	18	12	6	33.3%	A202018	15	3	12	80%
A212018	4	1	3	75%	A222018	16	4	12	75%
A232018	18	5	13	72.2%	A242018	10	8	2	20%
A252018	13	4	9	69.2%	A262018	4	0	4	100%
A272018	24	10	14	58.3%	A282018	13	9	4	30.8%
A292018	9	2	7	77.8%	A302018	11	4	7	63.6%
A312018	13	1	12	92.3%	A322018	21	9	12	57.1%
A332018	13	6	7	53.8%	B12018	19	7	12	63.2%
B152018	8	4	4	50%	B22018	19	10	9	47.4%
B162018	17	10	7	41.2%	B32018	24	14	10	41.7%
B172018	23	11	12	52.2%	B42018	14	11	3	21.4%
B182018	12	5	7	58.3%	B52018	12	8	4	33.3%
B192018	21	16	5	23.3%	B62018	15	5	10	66.7%
B202018	22	18	4	18.2%	B72018	25	8	17	68%
B212018	23	15	8	34.8%	B82018	18	8	10	55.6%
B222018	15	7	8	53.3%	B92018	20	9	11	55%
B232018	39	20	19	48.7%	B102018	15	8	7	46.7%
B242018	17	8	9	52.9%	B112018	21	4	17	81%
B252018	11	5	6	54.5%	B122018	16	12	4	25%
B262018	11	6	5	45.5%	B132018	25	15	10	40%
B272018	12	10	2	16.7%	B142018	16	7	9	56.3%
B282018	17	12	5	29.4%					
B292018	13	8	5	38.5%					
B302018	12	6	6	50%					
B312018	12	10	2	16.7%					
Average (all)	**15.15**	**7.8**	**7.3**	**50.7%**	**Average**	**15.3**	**7**	**8.3**	**53.3%**
Average (total 32)	**15.3**	**7.8**	**7.5**	**51.8%**					

Table 1 shows the data of Paragraph One improvement results, which were given feedback supposedly 50% out of 64 data with the LBI Bandscale and the other 50% with the ATA Framework. However, one of the raters misplaced some data so that the number of the

results assessed with the LBI Bandscale is a little bit higher (34) than that with the ATA Framework (30). To balance the number between the two, the alternative to calculate the average is by reducing two last data in red color (B302018 and B312018). Respondents starting with A in the code were from Class A and those with B were from Class B. The number of respondents in Class A is 33, while the number in Class B is 31, so they are in odd numbers and could not be divided evenly or precisely 50% while raters were assessing them.

From Table 1, it can be observed that the improvement results of Paragraph One with the LBI Bandscale range from 0% decrease (which means no improvement) to 92.3% decrease in errors. There is only one data showing no improvement. The improvement below 50% can be found in 14 data, while the improvement of 50% and more is found in 20 data. On average, the improvement of 34 data is 50.7%, but with 32 data, it is 51.8%. Meanwhile, the improvement results of Paragraph One with the ATA Framework range from 0% to 100% decrease in errors. There are 12 data showing the improvement of less than 50%, and 18 data show the improvement of more than 50%. Both the translation results of Paragraph One assessed with the LBI Bandscale and the ATA Framework each have more than 50% data showing improvement of more than 50%. This means that there is a great improvement of Paragraph One translation results in the revisions using both assessment models.

The next table contains the data of Paragraph Two improvement results.

Table 2. Paragraph Two Improvement Results

Respondent	The LBI Bandscale				Respondent	The ATA Framework			
	TE	RE	ED	ED%		TE	RE	ED	ED%
A22018	12	8	4	33.3%	A12018	10	6	4	40%
A42018	10	3	7	70%	A32018	13	1	12	92.3%
A62018	12	1	11	91.7%	A52018	14	6	8	57.1%
A82018	14	7	7	50%	A72018	14	3	11	78.6%
A102018	8	6	2	25%	A92018	12	2	10	83.3%
A122018	12	5	7	58.3%	A112018	11	1	10	90.9%
A142018	12	5	7	58.3%	A132018	15	6	9	60%
A162018	13	7	6	46.2%	A152018	8	6	2	25%
A182018	14	3	11	78.6%	A172018	12	6	6	50%
A202018	8	2	6	75%	A192018	16	10	6	37.5%
A222018	16	3	13	81.3%	A212018	10	3	7	70%
A242018	15	5	10	66.7%	A232018	16	6	10	62.5%
A262018	9	1	8	88.9%	A252018	13	3	10	76.9%
A282018	13	10	3	23.1%	A272018	28	16	12	42.9%
A302018	15	5	10	66.7%	A292018	14	4	10	71.4%
A322018	18	8	10	55.6%	A312018	14	10	4	28.6%
B12018	14	11	3	21.4%	A332018	14	7	7	50%
B22018	21	10	11	52.4%	B152018	17	7	10	58.8%
B32018	36	22	14	38.9%	B162018	15	12	3	20%
B42018	18	17	1	5.6%	B172018	24	10	14	58.3%
B52018	15	10	5	33.3%	B182018	20	3	17	85%
B62018	14	5	9	64.3%	B192018	31	21	10	32.3%
B72018	20	18	2	10%	B202018	29	15	14	48.3%
B82018	20	7	13	65%	B212018	33	10	23	69.4%
B92018	30	13	17	56.7%	B222018	24	13	11	45.8%
B102018	16	12	4	25%	B232018	33	20	13	39.4%
B112018	16	10	6	37.5%	B242018	19	6	13	68.4%
B122018	17	10	7	41.2%	B252018	12	6	6	50%
B132018	30	15	15	50%	B262018	18	7	11	61.1%
B142018	26	13	13	50%	B272018	19	8	11	57.9%
					B282018	23	16	7	30.4%
					B292018	14	5	9	64.3%
					B302018	17	8	9	52.9%
					B312018	21	10	11	52.4%

Average	**16.5**	**8.4**	**8.1**	**50.7%**	**Average (all)**	**17.7**	**8**	**9.7**	**56.2%**
					Average (total 32)	**17.7**	**8**	**9.7**	**56.5%**

The table shows that for Paragraph Two, there are 30 data assessed with the LBI Bandscale, while there are 34 data assessed with the ATA Framework. To make the number balance between the two assessment models, there is an alternative counting the average with only 32 data assessed with the ATA Framework. Nevertheless, the average of 34 data and that of 32 data are quite similar, 56.2% and 56.5%, respectively. The improvement results of Paragraph Two assessed with the LBI Bandscale range from 5.6% to 91.7%. There are 12 data showing improvement of less than 50%, whereas there are 18 data showing improvement of 50% and more. Similarly, the improvement of Paragraph Two assessed with the ATA Framework is also very substantial, as it ranges from 20% to 92.3%. Only 11 data show improvement of less than 50%, while there are 23 data of 50% improvement and above 50%.

From the findings, there is no doubt to believe that most participants have improved their translation results in the revisions greatly whether they received the LBI Bandscale or the ATA Framework as the assessment model that the raters used to assess their work. There is only one respondent showing no improvement in Paragraph One assessed with the LBI Bandscale, and there is another one having no improvement in Paragraph One assessed with the ATA Framework. The improvement in Paragraph One and Paragraph Two assessed with the LBI Bandscale of 50% and more than 50% is shown by more than 50% respondents. Similarly, the improvement in the paragraphs assessed with the ATA Framework show the same results. For the results assessed with the LBI Bandscale, the average of Paragraph One is 50.7% (34 data) and 51.8% (32 data), and the average of Paragraph Two is 50.7%. For the results assessed with the ATA Framework, the average of Paragraph One is 53.3%, and the average of Paragraph Two is 56.2% (34 data) and 56.5% (32 data). From the average, it can be seen that the results assessed with the ATA Framework show a higher percentage than those assessed with the LBI Bandscale. Does that mean that the ATA Framework is a more effective assessment model than the LBI Bandscale? Does the average percentage show a significant difference in the effectiveness? The calculation using *t*-test with SPSS in the following section will discover the answers.

The Effectiveness of the Assessment Models

The calculation using *t*-test is necessary to discover whether there is a significant difference in the effectiveness between the results assessed with the LBI Bandscale and those with the ATA Framework. The following is the definition of *t*-test from Heiman (2001).

> *The 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.* (Heiman, 2001, 393)

The *t*-test applied in this study is the independent sample *t*-test as there are two independent variables: the LBI Bandscale and the ATA Framework. The data entered in the *t*-test are from Tables 1 and 2 above.

The first data to be tested are the improvement results of Paragraph One as shown in Table 3 below.

Table 3. The Independent Sample *T*-Test for Paragraph One

Group Statistics

	ED	N	Mean	Std. Deviation	Std. Error Mean
Paragraph 1	ATA	30	7.7000	3.84304	.70164
	LBI	30	8.2667	4.35441	.79500

Independent Samples Test

	Levene's Test for Equality of		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
Paragraph 1 EVA	.930	.339	-.534	58	.595	-.56667	1.06034	-2.68918	1.55584
EVNA			-.534	57.118	.595	-.56667	1.06034	-2.68987	1.55654

Notes: EVA = Equal Variances Assumed; EVNA = Equal Variances Not Assumed

Before describing the results from Table 3 above, several terms need to be explained. The table shows that the total number of the respondents (N) is 60 (30+30), which can be tested for this *t*-test, so it is not 64. However, it is not N that determines the appropriate *t*-distribution for a study; it is df or degrees of freedom, and the formula of df is N – 2 (58) (Heiman, 2001, 375). The answer obtained from the *t*-test above is symbolized with t_{obt}, and the symbol for the critical (significant) value of *t* is t_{crit} (Heiman, 2001, 369), which indicates the boundary of the significant value of a sample mean. The result of t_{obt} is related to whether the hypothesis of this paper is accepted or rejected. In this section, the hypothesis is that there is a significant difference between the results assessed with the LBI Bandscale and those with the ATA Framework, and that those assessed with the ATA Framework show greater improvement; thus, this model is more effective than the LBI Bandscale. On the other hand, the null hypothesis ($H0$) which is "the statistical hypothesis that describes the population μs (the population mean) being represented if the predicted relationship does not exist" (Heiman, 2001, 363) of this section is that there is no significant difference of effectiveness between the LBI Bandscale and the ATA Framework as the assessment models in the improvement of the English-Indonesian translation results. According to Heiman, if t_{obt} is beyond t_{crit}, it means that the sample mean is in the region of rejection, and the null hypothesis can be rejected, while if t_{obt} is not beyond t_{crit}, it means that the sample mean does not lie in the region of rejection, and the null hypothesis cannot be rejected (2001, 370). The 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" (Heiman, 2001, 325).

Based on Table 3, the result of the t_{obt} is $t(58) = -.534$, $p > 0.05$, while the t_{crit} is ±2.000. The t_{crit} can be looked at Heiman (2001) page 708. The following figure will show the position of t_{obt}, whether it is or it is not beyond the t_{crit}.

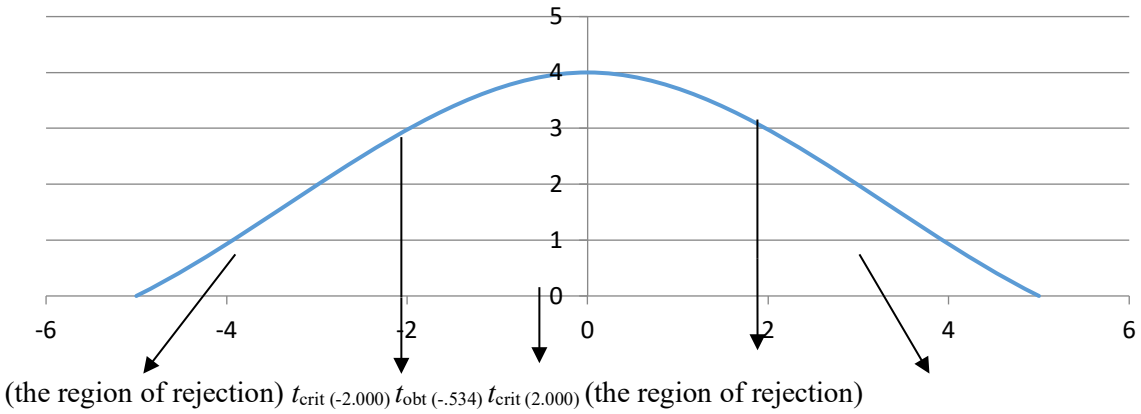

Figure 1. Sampling Distribution of Paragraph One

From the figure above, it is obvious that the t_{obt} (-.534) does not lie in the region of rejection. It means the null hypothesis cannot be rejected, or the hypothesis is rejected. In other words, the improvement results of Paragraph One between the ones assessed with the LBI Bandscale and those with the ATA Framework do not indicate any significant difference in effectiveness. It means both assessment models have the same effectiveness, at least in the improvement results of Paragraph One.

The following table shows the results for Paragraph Two. The result the t_{obt} is $t(58) = -1.582$, $p > 0.05$, while the t_{crit} is ±2.000. This t_{obt} (-1.582) does not lie in the region of rejection, either, as it can be seen in Figure 2. Thus, it means there is no significant difference in the effectiveness between the LBI Bandscale and the ATA Framework in the improvement of Paragraph Two. Although based on the average percentage the ATA Framework (56%) show a higher percentage than the LBI Bandscale (50.7%), the difference is considered not significant according to the *t*-test results.

Table 4. The Independent Sample *T*-Test for Paragraph Two

Group Statistics

	ED	N	Mean	Std. Deviation	Std. Error Mean
Paragraph 2	ATA	30	8.0667	4.19304	.76554
	LBI	30	9.8000	4.29434	.78404

Independent Samples Test

	Levene's Test for Equality of		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
Paragraph 2 EVA	.385	.537	-1.582	58	.119	-1.73333	1.09579	-3.92680	.46014
Paragraph 2 EVNA			-1.582	57.967	.119	-1.73333	1.09579	-3.92683	.46016

Notes: EVA = Equal Variances Assumed; EVNA = Equal Variances Not Assumed

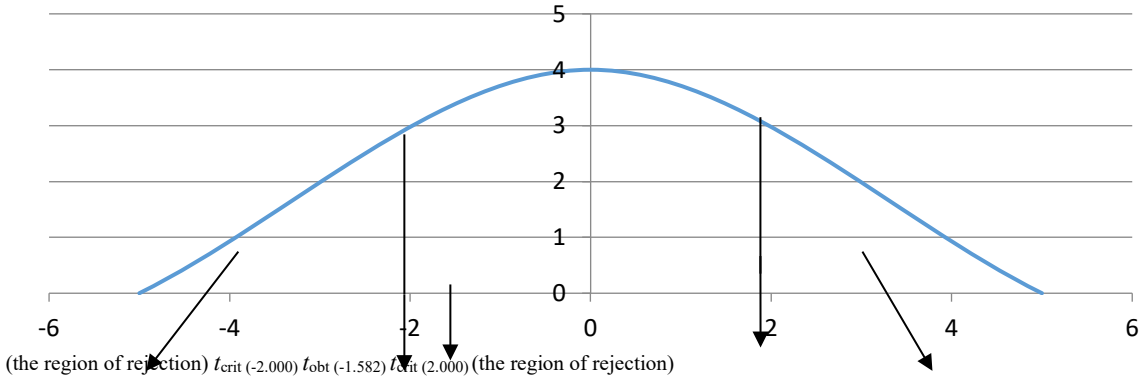

Figure 2. Sampling Distribution of Paragraph Two

The findings above show that the improvement of translation results in both paragraphs One and Two, using either the LBI Bandscale or the ATA Framework as the assessment models, is quite similar. Based on the results of the *t*-test conducted, it is evident there is no significant difference in the effectiveness of the two models applied. Conde (2009, 2011, 2012a, 2012b) and Turner et al. have discovered the same findings when they researched on the effectiveness between the holistic approach assessment and the analytical approach assessment, but they did not use the LBI Bandscale (the holistic approach) and the ATA Framework (the analytical approach). Dewi (2015) applied those two rubrics as used in this present study and discovered that there is no significant difference in the effectiveness of the two models for the Indonesian to English translation results, but the number of the participants involved was just half the number of the participants in this study. Thus, Dewi's work (2015) is considered as a preliminary study since the number of the research subjects is not sufficient to lead to conclusive evidence, even though the data collection was conducted for two years. This present research is the continuity of Dewi's work (2015) in a different translation direction, from English to Indonesian, with a significant number of the participants, to provide further evidence that whether a translation result is assessed by the holistic approach or by the analytical approach, its improvement will be more or less similar.

Conclusion

The findings of this study show that the participants had great improvement in their translation results in the revisions whether they received the LBI Bandscale or the ATA Framework as the assessment model which was applied to assess their works. Despite the fact that the average percentage of the results assessed with the ATA Framework is a little bit higher than the average percentage of those with the LBI Bandscale, the *t*-test results indicate that there is no significant difference in the effectiveness between the two assessment models both in Paragraph One and in Paragraph Two. These findings confirm Conde's (2009, 2011, 2012a, 2012b), Turner et al.'s (2010), and Dewi's (2015) works that the holistic system assessment and the analytical system assessment are both effective or have similar effectiveness. Although this type of study has been conducted by several Translation Studies scholars with similar findings, only Dewi (2015) and this present research focus on the Indonesian and English language pair. In spite of the same results, we cannot make an assumption that the findings from a certain language pair translation can be applied into another language pair translation without any research conducted to support it.

The research findings are limited to the population of novice translators, as the participants involved were students, and the text used to be translated was only one kind, which is an academic text. For future research, these two types of assessment models need to

be combined and tested to discover the effectiveness. In addition, the research can focus on different types of source texts, such as journalistic texts, legal texts, manual texts, and others, to be translated and assessed. The participants or the respondents can come from the graduate classes or the Master's degree program, or they can be professional translators, so there will be new insight and findings from different research populations. It is expected that this study can encourage more research in translation assessment and the establishment of an appropriate assessment model for English to Indonesian translation.

Acknowledgments

The authors are grateful for the 2018 Universitas Indonesia PITTA grant provided to complete this paper. We are also thankful for the respondents participated in this study.

References

Arango-Keeth, F. & Koby, G. S. (2003) Assessing Assessment: Translator Training Evaluation and the Needs of Industry Quality Assessment. In B. J. Baer & G. S. Koby (Eds.). *Beyond the Ivory Tower: Rethinking Translation Pedagogy (ATA Scholarly Monograph Series Volume XII)* (pp. 117-134). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Conde, T. (2009) *Proceso y Resultado de la Evaluación de Traducciones* [The Process and Results of Translation Evaluation]. Granada: Universidad de Granada, PhD Dissertation.

_____. (2011) Translation Evaluation on the Surface of Texts: A Preliminary Analysis. *JoSTrans, 15*, 69-86.

_____. (2012a) Quality and Quantity in Translation Evaluation: A Starting Point. *Across Languages and Cultures, 13*(1), 67-80.

_____. (2012b) The Good Guys and the Bad Guys: The Behavior of Lenient and Demanding Translation Evaluators. *Meta, 57*(3), 763-786.

_____. (2013) Translation versus Language Errors in Translation Evaluation. In D. Tsagari & R. van Deemter (eds.), *Assessment Issues in Language Translation and Interpreting*. Frankfurt am Main: Peter Lang Publishing, 97-112.

Cravo A. & Neves, J. (2007) Action Research in Translation Studies. *The Journal of Specialised Translation*, 7, 92–107.

Dewi, H. D. (2015) *Comparing Two Translation Assessment Models: Correlating Student Revisions and Perspectives*. (Unpublished Ph.D. Dissertation. Kent State University, Ohio.

Dick, B. (1993) "You Want to Do an Action Research Thesis?" Online at http://www.scu.edu.au/schools/gcm/ar/art/arthesis.html (consulted on 4 July 2020).

Doyle, M. S. (2003) Translation Pedagogy and Assessment: Adopting ATA's Framework for Standard Error Marking. *The ATA Chronicle, 32*(11), 21-28 & 45.

Heiman, G. W. (2001) *Understanding Research Methods and Statistics*. Boston, MA: Houghton Mifflin Company.

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.

Qu, W. & Zhang, C. (2013) The Analysis of Summative Assessment and Formative Assessment and their Roles in College English Assessment System. *Journal of Language Teaching and Research, 4*(2), 335-339.

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.

Vinay, J. & Darbelnet, J. (1995) *Comparative Stylistics of French and English.* Amsterdam & Philadelphia: John Benjamins Publishing Company.

Waddington, C. (2001a) Should Translations be Assessed Holistically or through Error Analysis? *Hermes, Journal of Linguistics*, *26*, 15-37.

_____. (2001b) Different Methods of Evaluating Student Translations: The Question of Validity. *Meta, 46*(2), 311-325.

_____. (2003) A Positive Approach to the Assessment of Translation Errors. In R. Muñoz-Martín (Ed.), *AIETI, 2,* 409-426. Granada: AIETI.

APPENDIX 1

THE LBI BANDSCALE
ASSESSMENT BANDSCALE FOR PPP LBI UI TRANSLATOR TRAININGS

	Indonesian	**English**	
A	Pemahaman TSu dan penulisan TSa baik sekali. Sesekali secara kreatif mampu menemukan padanan yang sangat sesuai.	The source text (ST) understanding and the target text (TT) writing are excellent. Sometimes the student can creatively discover very suitable equivalents.	A = 4 A- = 3.5
B	Pemahaman TSu baik, namun adakalanya terjadi kesalahpahaman TSu, terutama jika menerjemahkan bagian teks yang sulit. Penulisan dalam BSa umumnya baik, tidak banyak membuat kesalahan ejaan dan/atau tanda baca.	The ST understanding is good, but occasionally there is ST misunderstanding, especially when translating a difficult part of the text. The writing in the target language (TL) is generally good, not making many errors in spelling and/or punctuation marks.	B+ = 3.2 B = 3 B- = 2.8
C	Pemahaman TSu cukup baik jika tingkat kesulitan teks tidak tinggi. Namun jika teks memiliki banyak ungkapan idiomatis atau terminologi khusus, peserta sering tidak mampu memahami teks dengan baik. Dalam hal penulisan dalam Bsa, peserta seringkali membuat kesalahan yang terkait dengan pilihan kata, kolokasi, ejaan dan tanda baca.	The ST understanding is quite good if the text difficulty level is not high. However, if the text has a large number of idiomatic expressions or special terminology, the student will often be unable to understand the text well. In terms of the TL writing, the student often makes mistakes related to choices of words, collocation, spelling, and punctuation marks.	C+ = 2.5 C = 2 C- = 1.5
D	Pemahaman TSu perlu ditingkatkan lagi. Banyak kesalahan pengungkapan pesan ke dalam Bsa yang menyebabkan salah pengertian.	The ST understanding needs to be improved further. Many errors in the message transfer into the TL causing misunderstanding.	D = 1

APPENDIX 2

THE ATA FRAMEWORK

Exam No.: _____

Passage No.: _____

ATA CERTIFICATION PROGRAM
FRAMEWORK FOR STANDARDIZED ERROR MARKING
Version 2017

1	2	4	8	16	Code	Reason
Errors that concern the form of the exam						
Treat missing material within the passage as an omission.					UNF	Unfinished (If a passage is substantially unfinished, do not grade the exam.)
					ILL	Illegibility
					IND	Indecision, gave more than one option
Meaning transfer or strategic errors: Negative impact on clarity or usefulness of target text. Use one of the categories below whenever possible. If none are applicable, use OTH-MT						
					A	Addition
					AMB	Ambiguity
					COH	Cohesion
					F	Faithfulness (translation strays too far from ST meaning)
					FA	Faux ami (false friend)
					L	Literalness
					MU	Misunderstanding of source text (if identifiable)
					O	Omission
					T	Terminology, word choice
					TT	Text type (failure to follow Translation Instructions) This category also covers register and style.
					VT	Verb Tense (grammar correct, but conveys wrong meaning)
					OTH-MT	Other (describe; use a separate page if needed)
Mechanical errors: Negative impact on overall quality of target text. Points may vary by language. Maximum 4 points. **Use one of the categories below whenever possible. If none are applicable, use OTH-ME**						
					G	Grammar (use one of next two sub-categories if applicable)
					SYN	— Syntax (phrase/clause/sentence structure)
					WF/ PS	—Word form /Part of speech
					P	Punctuation
					SP/ CH	Spelling/Character (usually 1 point, maximum 2; if more than 2 points, another category must apply)
					D	— Diacritical marks/Accents
					C	— Capitalization
					U	Usage
					OTH-ME	Other (describe; use a separate page if needed)

	x 2 =	x 4 =	x 8 =	x 16 =		**Column totals**
A grader may stop marking errors when score reaches 46 error points (mark such exams 46+)			A grader may award a quality point for each of up to three instances of exceptional translation.			Quality points are subtracted from the error point total to yield a final score. A passage with a score of 18 or more points receives a grade of Fail.
Total error points (add column totals):			**Quality points (maximum 3):**			**Final passage score (subtract quality points from error points):**

APPENDIX 3

THE CONSENT FORM
Informed Consent to Participate in a Research Study

Study Title:
The effectiveness between two translation assessment models for English to Indonesian translation
Principal Investigator: Haru Deliana Dewi, Ph.D.
Co-Investigator: Prof. Dr. Rahayu Surtiati Hidayat
You are being invited to participate in a research study. This consent form will provide you with information on the research project, what you will need to do, and the associated risks and benefits of the research. Your participation is entirely voluntary. Please, read this form carefully. It is important that you ask questions and fully understand the research in order to make an informed decision. You will receive a copy of this document to take with you.
Purpose: This study aims to discover the effectiveness of two translation assessment models to improve translation results.
Procedure: You will be presented with a short source text (2 paragraphs, each consisting of around 175 to 200 words) and be requested to translate them into Indonesian in class. You will be expected to work individually for one (1) hours and can use any resources available. After you have finished, please send the results to harudd.dewi7@gmail.com. Two weeks later, you will receive the assessment and feedback on your work. One paragraph will be assessed using one model of assessment, and the other paragraph will be assessed using the other model of assessment. Next, you will be requested to revise your first versions based on this assessment and feedback. You will do the revision in class for one hour. Please also send the revisions to harudd.dewi7@gmail.com. Then you will be asked to fill in a simple short survey to report your perspective on these two models of assessment.
Benefits: The potential benefits of participating in this study include having your translation assessed and receiving feedback to help improve your translation skills. Once you have completed your participation, you will be entitled to receive a token of appreciation, such as a keychain or a pen or something similar.
Risks and Discomforts: There are no anticipated risks beyond those encountered in everyday life.
Privacy and Confidentiality: No identifying information will be collected. Your signed consent form will be kept separate from your study data, and responses/results will not be linked to you.
Voluntary Participation: Taking part in this research study is entirely voluntary. You may choose not to participate or you may discontinue your participation at any time without penalty or loss of benefits to which you are otherwise entitled. You will be informed of any new, relevant information that may affect your health, welfare, or willingness to continue your study participation.
Contact Information: If you have any questions or concerns about this research, you may contact Haru Deliana Dewi at harudd.dewi7@gmail.com.

(The accompanying Indonesian form is a true and fair equivalent of this original English form. See the third and fourth pages of the Indonesian version of this document.)
I have read this consent form and have had the opportunity to have my questions answered to my satisfaction. I voluntarily agree to participate in this study. I understand that a copy of this consent will be provided to me for future reference.

(*Saya telah membaca formulir persetujuan tertulis ini dan pertanyaan saya telah terjawab secara memuaskan. Saya secara sukarela setuju untuk berpartisipasi dalam penelitian ini. Saya mengerti bahwa salinan persetujuan tertulis ini akan disediakan bagi saya untuk rujukan di masa depan.*)

_____ _____
Participant Signature Date
(*Tanda Tangan Peserta*) (*Tanggal*)

APPENDIX 4

INSTRUCTIONS AND THE SOURCE TEXT

INSTRUCTIONS FOR THE FIRST ASSIGNMENT:
TRANSLATION FROM ENGLISH TO INDONESIAN
(*INSTRUKSI UNTUK TUGAS PERTAMA:*
***PENERJEMAHAN DARI BAHASA INGGRIS KE BAHASA INDONESIA*)**

Instructions for the first assignment (*instruksi untuk tugas pertama*):
Translation of a short general text from English to Indonesian (*Penerjemahan teks umum pendek dari bahasa Inggris ke bahasa Indonesia*)

- Your task will be to translate the colored highlighted part of the following document in the space provided below the source text. (*Tugas anda adalah menerjemahkan bagian berwarna dari dokumen di bawah ini dan menerjemahkannya di bawah teks sumber.*)
- You may have only one hour to complete the task. (*Anda dapat menyelesaikan tugas ini hanya selama satu jam.*)
- You may use any resources, such as dictionaries, internet, and others. (*Anda diperbolehkan menggunakan segala alat bantu penerjemahan, seperti kamus, internet, dan lain-lain.*)
- Please work individually. (*Mohon bekerja sendiri-sendiri.*)
- Do not consult anyone else for help. (*Jangan bekerja sama dengan orang lain dalam menerjemahkan.*)
- Please name your file with your first name and the type of assignment. (*Mohon menamai berkas anda ini berdasarkan nama pertama anda dan tipe tugas.*)
- For example, if your name is Anton Fermana and you are doing a translation, then the file name will be antonfermana_translation.doc. (*Misalnya, jika nama anda Anton Fermana dan anda mengerjakan penerjemahan, maka nama berkas ini menjadi antonfermana_translation.doc.*)
- After you have finished, please submit the work to harudd.dewi7@gmail.com. (*Setelah anda selesai, mohon kirim hasilnya ke harudd.dewi7@gmail.com.*)

English source text: (*Teks Sumber Berbahasa Inggris:*)

The Creators of Grammar

No student of a foreign language needs to be told that grammar is complex. By changing word sequences and by adding a range of auxiliary verbs and suffixes, we are able to communicate tiny variations in meaning. We can turn a statement into a question, state whether an action has taken place or is soon to take place, and perform many other word tricks to convey subtle differences in meaning. Nor is this complexity inherent to the English language. All languages, even those of so-called 'primitive' tribes have clever grammatical components. The Cherokee pronoun system, for example, can distinguish between 'you and I', 'several other people and I' and 'you, another person

and I'. In English, all these meanings are summed up in the one, crude pronoun 'we'. Grammar is universal and plays a part in every language, no matter how widespread it is. Thus, the question which has baffled many linguists is - who created grammar?

At first, it would appear that this question is impossible to answer. To find out how grammar is created, someone needs to be present at the time of a language's creation, documenting its emergence. Many historical linguists are able to trace modern complex languages back to earlier languages, but in order to answer the question of how complex languages are actually *formed*, the researcher needs to observe how languages are started from scratch. Amazingly, however, this is possible.

Some of the most recent languages evolved due to the Atlantic slave trade. At that time, slaves from a number of different ethnicities were forced to work together under colonizer's rule. Since they had no opportunity to learn each other's languages, they developed a make-shift language called a *pidgin*. Pidgins are strings of words copied from the language of the landowner. They have little in the way of grammar, and in many cases it is difficult for a listener to deduce when an event happened, and who did what to whom. Speakers need to use circumlocution in order to make their meaning understood. Interestingly, however, all it takes for a pidgin to become a complex language is for a group of children to be exposed to it at the time when they learn their mother tongue. Slave children did not simply copy the strings of words uttered by their elders, they adapted their words to create a new, expressive language. Complex grammar systems which emerge from pidgins are termed creoles, and they are invented by children.

Further evidence of this can be seen in studying sign languages for the deaf. Sign languages are not simply a series of gestures; they utilize the same grammatical machinery that is found in spoken languages. Moreover, there are many different languages used worldwide. The creation of one such language was documented quite recently in Nicaragua. Previously, all deaf people were isolated from each other, but in 1979 a new government introduced schools for the deaf. Although children were taught speech and lip reading in the classroom, in the playgrounds they began to invent their own sign system, using the gestures that they used at home. It was basically a pidgin. Each child used the signs differently, and there was no consistent grammar. However, children who joined the school later, when this inventive sign system was already around, developed a quite different sign language. Although it was based on the signs of the older children, the younger children's language was more fluid and compact, and it utilized a large range of grammatical devices to clarify meaning. What is more, all the children used the signs in the same way. A new creole was born.

Some linguists believe that many of the world's most established languages were creoles at first. The English past tense –ed ending may have evolved from the verb 'do'. 'It ended' may once have been 'It end-did'. Therefore, it would appear that even the most widespread languages were partly created by children. Children appear to have innate grammatical machinery in their brains, which springs to life when they are first trying to make sense of the world around them. Their minds can serve to create logical, complex structures, even when there is no grammar present for them to copy.

INSTRUCTIONS FOR THE REVISION
(*INSTRUKSI UNTUK REVISI*)

Instructions for the revision: (*instruksi untuk revisi:*)

- Please read your assessment and feedback. Study the assessment carefully. (*Mohon baca penilaian dan umpan balik anda. Pelajari penilaiannya secara seksama.*)
- Please revise your initial assignment based on the assessment and feedback given to you. (*Mohon revisi tugas pertama anda berdasarkan penilaian dan umpan balik yang diberikan kepada anda.*)
- Copy your first assignment in the space provided below to be revised there. (*Salin tugas pertama anda pada tempat yang telah disediakan di bawah ini untuk direvisi.*)
- Please work individually. (*Mohon bekerja sendiri-sendiri.*)
- Do not consult anyone else for help. (*Jangan bekerja sama dengan orang lain dalam merevisi tugas anda.*)
- You will have one hour to complete the task. (*Anda akan mempunyai waktu satu jam untuk menyelesaikan tugas ini.*)
- Name your file correctly! (*Beri nama berkas Anda dengan benar.*)
- For example, if your name is Anton Fermana and you are doing a revision, then the file name will be antonfermana_revision.doc. (*Misalnya, jika nama anda Anton Fermana dan anda mengerjakan revisi, maka nama berkas ini menjadi antonfermana_revision.doc.*)
- After you have finished, please send the revision to harudd.dewi7@gmail.com. (*Setelah anda selesai, mohon kirim revisi anda ke harudd.dewi7@gmail.com.*)
- When you have sent the revision, please go https://www.surveymonkey.com/r/GJXWK9P to take a short survey. (*Ketika anda telah mengirimkan revisi anda, mohon mengisi survei pendek dengan mengklik situs berikut ini https://www.surveymonkey.com/r/GJXWK9P.*)
- Thank you very much! (*Terima kasih banyak.*)