

**MEANING-ORIENTED ASSESSMENT: AN ANALYSIS
OF ITS INTER-RATER RELIABILITY, STRENGTHS,
AND WEAKNESSES**

SUWARNI WIJAYA HALIM

2014-001-004

**A thesis submitted in partial fulfillment of the requirements for the Degree of
Master of Arts in Applied English Linguistics**



**APPLIED ENGLISH LINGUISTICS PROGRAM
FACULTY OF EDUCATION AND LANGUAGE
ATMA JAYA CATHOLIC UNIVERSITY OF INDONESIA
2017**

THESIS APPROVAL FORM

Name : Suwarni Wijaya Halim
Student Number : 2014001004
Program : Magister in Applied English Linguistics
Major : Applied English Linguistics
Date of Examination : 4th May 2017
Thesis Title : Meaning-Oriented Assessment: An Analysis on its Inter-Rater Reliability, Strengths, and Weaknesses

Examination Board

Prof. Bahren Umar Siregar, Ph.D.
Head of MA Program in Applied
English Linguistics

Prof. Bahren Umar Siregar, Ph.D.
Thesis Advisor

Yassir Nasanius, Ph.D.
Examiner 1

Dr. Setiono Sugiharto
Examiner 2

STATEMENT OF WORK ORIGINALITY

I hereby declare that this thesis is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of materials which have been accepted for the award of any other educational institutions, except where due acknowledgement is made in the Thesis.

Jakarta, 9th May 2017

Suwarni Wijaya Halim

ABSTRACT

This research looked into the implementation of Kim's model of meaning-oriented assessment (2009) in assessing students' translation works. The motivation behind this topic is the fact that there has been lack of research regarding the topic of translation assessment in the area of translation pedagogy, and there has been almost no previous research about the usage of meaning-oriented assessment in Indonesian context. This research attempted to fill the gap and aims at discovering the result of assessment using meaning-oriented assessment model, determining the inter-reliability of the model, and finding out the strengths and weaknesses of the model. To achieve the aims of the research, the researcher utilized mixed method with triangulation as the research design. Selected students were asked to translate an editorial taken from newspaper from English to Indonesian, and the results of their translation were assessed by three raters using meaning-oriented assessment. The assessment sheets were observed and compared, and the scores given by the raters were inputted into SPSS in order to prove whether or not the model has inter-rater reliability. Afterwards, interview towards the raters were conducted in order to find out the strengths and weaknesses of the model. The result of the analysis shows that meaning-oriented assessment either had low level of inter-rater reliability or no inter-rater reliability, and according to the results of interview, the researcher found that the raters perceived the detailed criteria and the high validity as the strengths of meaning oriented assessment. Nevertheless, they also thought that meaning-oriented assessment is too time-consuming and complicated. In addition, they believed meaning-oriented assessment might not produce reliable results, and its implementation might be problematic, both for formative and summative assessment. As a final remark, the researcher believed that the implementation of meaning-oriented assessment would require great effort from the translator trainers, so the researcher suggested modification or proper and sufficient training process for any parties who would like to implement meaning-oriented assessment.

Keywords: *meaning-oriented translation assessment, translation pedagogy*

ACKNOWLEDGEMENT

I would like to convey my deepest gratitude to Lord Buddha. With His blessings, I finally managed to finish this thesis. My next deepest gratitude goes to my family; my late grandmothers, my mother, my father, and my younger brother. They never stop motivating me throughout my research process, and they always remind me to never stop writing. Despite the fact that I took so much time in finishing my thesis, I am really happy that they are always there and constantly pray for me.

I would also like to thank my thesis advisor, Professor Bahren Umar Siregar, Ph.D. Thank you very much for your time, inputs, and guidance. Despite your hectic schedules, you constantly give me suggestions so that I can develop and expand my thesis further. I really appreciate everything that you have done for me. My gratitude also goes to Mr. Agustian Sutrisno, Ph.D. Thank you for directing me during the interview process and for providing me with valuable suggestions. I am aware that there are a lot of shortcomings in this thesis, but I hope that my research will meet your expectation. I also would like to express my gratitude to Dr. Mira Kim. Your research has given me the idea for my thesis, and your replies to my emails have encouraged me to continue working on it.

Also, I would like to thank Frans Seda Foundation and Dr. Menno van Zaanen. To Frans Seda Foundation, thank you for granting me the opportunity to see the new world and to meet amazing people in the Netherlands. I was able to gain new perspectives and points of view for my research; all because of Frans Seda

Foundation. To Menno, you are a wonderful advisor. Thank you for giving me new ideas for my research. Unfortunately, I could not implement your ideas into this thesis. Hopefully there will be another opportunity where I can apply those ideas.

I am also grateful for the supports provided by my lovely colleagues in Bunda Mulia University: Ibu Murniati, Bapak Ardi Nugroho, Bapak Alvin Taufik, Bapak Ronald Silalahi, Ibu Nurdiana, Ibu Magdalena Kartikasari, Bapak Jonathan Tanihardjo, Bapak Rex Stardy, and Ibu Maria Tamarina Prawati. Thank you for constantly supporting me throughout my research process. Taking up higher education while working is not a walk in the park, but you have helped me a lot and made everything easier for me. I promise that I will help you with anything once my thesis process is officially concluded.

Next, I would like to express my gratitude to the raters who are involved in this research. Thank you very much for your time and your willingness to participate in my “crazy” research. I apologize if I gave you hard time during my research process. Without you and your participation, this research would not have been completed.

I would also like to thank my dearest friends in Master class batch 2014, particularly Hilda Novita and Christela Cindy. When I was confused or feeling down, you never hesitated to provide insights and assistance. I cannot thank you enough for everything you have done for me. Let’s have dinner together sometimes to celebrate the completion of our thesis.

My next gratitude goes to my best friends, Robby Wijaya dan Devi Wijaya. You brighten up my days, and I would never forget that. Whenever I am at my lowest

point, you are always there to cheer me up and lift my spirits. You are truly the best friends that I have ever had.

Then, to all people whose name I cannot mention one by one, thank you very much for your support. Without all of you, I might not be able to complete this thesis. Finally, as a concluding remark, to all people I have mentioned and have not mentioned in this acknowledgement, I dedicate this thesis to you.

TABLE OF CONTENTS

| | |
|-------------------------------------|------|
| COVER | i |
| THESIS APPROVAL FORM | ii |
| STATEMENT OF WORK ORIGINALITY | iii |
| ABSTRACT | iv |
| ACKNOWLEDGMENT | v |
| TABLE OF CONTENTS | viii |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xii |

CHAPTER 1 : INTRODUCTION

| | |
|------------------------------------|---|
| 1.1. Background of Research | 1 |
| 1.2. Research Questions | 6 |
| 1.3. Research Objectives | 6 |
| 1.4. Research Significance | 7 |
| 1.5. Scope of Research | 7 |
| 1.6. Limitation of Research | 8 |
| 1.7. Organization of Writing | 9 |

CHAPTER 2 : LITERATURE REVIEW

| | |
|---|----|
| 2.1. Translation Assessment in Translator Training | 10 |
| 2.2. Systemic Functional Linguistics | 14 |
| 2.3. Meaning-Oriented Assessment | 17 |
| 2.3.1. Rationale and Implementation of Meaning-Oriented Assessment | 18 |
| 2.3.2. Example of Error Analysis in Meaning-Oriented Assessment | 21 |
| 2.4. Criteria of Assessment Tool | 22 |
| 2.5. Previous Studies on Translation Assessment in Indonesian Context | 24 |
| 2.6. Theoretical Framework | 25 |

CHAPTER 3 : RESEARCH METHODOLOGY

| | |
|------------------------------|----|
| 3.1. Research Paradigm | 26 |
| 3.2. Research Design | 27 |
| 3.3. Validity Issue | 28 |
| 3.4. Data Sources | 28 |

| | |
|---|----|
| 3.4.1. Selection of Research Participants | 29 |
| 3.4.2. Selection of Translation Task | 29 |
| 3.4.3. Selection of Raters | 31 |
| 3.4.4. Pilot Study | 32 |
| 3.4.5. Guide to Meaning-Oriented Assessment | 33 |
| 3.4.6. Interview Protocol & Questions | 34 |
| 3.6. Data Collection and Analysis | 36 |

CHAPTER 4 : FINDINGS & DISCUSSION

| | |
|--|----|
| 4.1. Comparison of Results of Assessment by Rater | 39 |
| 4.1.1. Similar Error Categories, Similar Point Deduction | 39 |
| 4.1.2. Similar Error Categories, Different Point Deduction | 43 |
| 4.1.3. Different Error Categories, Different Point Deduction | 48 |
| 4.1.4. Different Overall Assessment Results | 49 |
| 4.2. Inter-Rater Reliability | 55 |
| 4.3. Raters' Perception on Meaning-Oriented Assessment | 58 |
| 4.3.1. Strengths of Meaning-Oriented Assessment | 59 |
| 4.3.1.1. Detailed Criteria of Assessment | 59 |
| 4.3.1.2. High Validity | 61 |
| 4.3.2. Weaknesses of Meaning-Oriented Assessment | 63 |
| 4.3.2.1. Time Inefficiency | 63 |
| 4.3.2.2. Complicated Mechanism | 65 |
| 4.3.2.3. Subjective Interference | 67 |
| 4.3.2.4. Low Reliability | 69 |
| 4.3.2.5. Problematic Implementation | 72 |
| 4.4. Result of Triangulation | 74 |

CHAPTER 5 : CONCLUSION & SUGGESTIONS

| | |
|------------------------|----|
| 5.1. Conclusion | 77 |
| 5.2. Suggestions | 80 |

| | |
|-------------------------|----|
| REFERENCES | 81 |
|-------------------------|----|

LIST OF APPENDICES

| | |
|--|----|
| APPENDIX 1 : SOURCE TEXT | 86 |
| APPENDIX 2 : INTERVIEW PROTOCOL | 87 |
| APPENDIX 3 : WRITTEN CONSENT FORM (for participants) | 90 |

APPENDIX 4 : RESEARCH BACKGROUND (for raters) 91
APPENDIX 5 : WRITTEN CONSENT FORM (for raters) 93
APPENDIX 6 : SAMPLE OF ASSESSMENT SHEET (Student 1) 94
APPENDIX 7 : GUIDE TO MEANING-ORIENTED ASSESSMENT 102

LIST OF TABLES

| | |
|---|----|
| Table 1. Comparison of Assessment Results (Sentence 1; Student 6) | 40 |
| Table 2. Comparison of Assessment Results (Sentence 2; Student 5) | 40 |
| Table 3. Comparison of Assessment Results (Sentence 12; Student 10)..... | 41 |
| Table 4. Comparison of Assessment Results (Sentence 7; Student 3) | 41 |
| Table 5. Comparison of Assessment Results (Sentence 10; Student 5) | 42 |
| Table 6. Comparison of Assessment Results (Sentence 1; Student 5) | 44 |
| Table 7. Comparison of Assessment Results (Sentence 1; Student 10) | 44 |
| Table 8. Comparison of Assessment Results (Sentence 3; Student 2) | 45 |
| Table 9. Comparison of Assessment Results (Sentence 4; Student 10) | 47 |
| Table 10. Comparison of Assessment Results (Sentence 20; Student 10)..... | 48 |
| Table 11. Comparison of Assessment Results (Sentence 2; Student 1) | 49 |
| Table 12. Comparison of Assessment Results (Sentence 1; Student 1) | 50 |
| Table 13. Comparison of Assessment Results (Sentence 2; Student 3) | 50 |
| Table 14. Comparison of Assessment Results (Sentence 2; Student 8) | 51 |
| Table 15. Comparison of Assessment Results (Sentence 12; Student 1) | 52 |
| Table 16. Comparison of Assessment Results (Sentence 11; Student 9) | 52 |
| Table 17. Summary of Point Deduction | 55 |
| Table 18. Summary of Final Scores | 56 |
| Table 19. Result of Normality Test | 57 |
| Table 20. Result of Kruskal-Wallis Test | 58 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1. Stratification of Language | 15 |
| Figure 2. Relationship between Grammar (Lexicogrammar), Semantics (Metafunctions) and Context (Register)..... | 15 |
| Figure 3. Modalization and Modulation | 17 |
| Figure 4. Relation among Register, Metafunction, and Lexicogrammatical Elements..... | 18 |
| Figure 5. Criteria for Meaning-Oriented Assessment | 19 |
| Figure 6. Theoretical Framework | 25 |
| Figure 7. Illustration of Triangulation Process | 74 |

CHAPTER 1

INTRODUCTION

1.1. Background of Research

Despite the fact that the activities of translation had been conducted for a long time, Translation Studies is considered as a discipline which is relatively ‘new’ or ‘young’ in the research community. Firstly and officially outlined by Holmes in 1988 (as cited in Munday, 2008), Translation Studies can be divided into two major areas: ‘pure’ and ‘applied’ studies. ‘Pure’ areas of translation studies attempt to describe the phenomena in translation activities and establish the theories of translation based on those phenomena. ‘Applied’ areas, on the other hand, are more concerned with the practical domain of translation activities, and the framework is classified further into three: translator training, translator aid, and translation criticism. This research focused dominantly on the area of ‘applied’ studies, particularly aspect of translation criticism in translator training.

Translation criticism or translation assessment or translation evaluation (hereinafter, translation assessment), as the name suggests, involves assessing or evaluating the value or the quality of a translation. Since quality is an important element in translation, it is imperative to conduct empirical research about translation assessment in order to form the most appropriate assessment models to acquire a conclusion about translation quality (Saldanha & O’Brien, 2014). Despite this logic, however, the topic of translation assessment is still under-researched.

Numerous experts in translation studies actually consider translation assessment as a robust area to study and develop, yet there are very few and limited resources, literatures, and previous studies (Hatim & Mason, 1997; Angelelli & Jacobson, 2009). One of the possible reasons for the lack of previous studies is because translation assessment is a very complex topic of research, and it requires elaborated and detailed procedures in the implementation (Saldanha & O'Brien, 2014). Melis and Albir (2001) also explained that translation assessment is rarely conducted, and if it is conducted at all, subjective interpretations and criteria often dominate the process of assessment. This notion is also supported by Colina (2003) who stated that translation assessment is often done in unclear manner and that there have been too few literatures which discuss about the proper application and implementation of translation assessment.

Melis and Albir (2001) classified translation assessment into three different categories. The first category is the assessment for published translation. In this case, the objects of assessment are usually literary texts (such as novels, prose, poems, and so on) and religious texts (such as the Koran, the Bible, and so on). The main goal of this assessment is “to judge a translation, to discuss its merits and demerits, and, sometimes, to propose solutions” (Melis & Albir, 2001, p. 273). The second category is the assessment for professional translators. In this case, the focus of the assessment is the translators, along with the translation competence that they possess. This kind of assessment is usually conducted to certify the translators, to test the translators’ abilities before joining a certain professional organization, and so on. The types of texts used in this assessment vary from literary texts to legal texts since translators are

expected to be proficient in translating any kinds of texts. The last category is the assessment for translator training. This type of assessment involves identifying, correcting, and weighing the value of errors and giving scores or grades. The focus of the assessment is, of course, translation assignments or projects by translation students. Each of those categories demands different model of translation assessment.

In the case of translator training, the notion of translation assessment becomes very important. It is a way for the students to discover how they perform, which errors they often make, and which parts they need to improve on. It is also a way for the translator trainers to find out whether the students have improved, which parts they need to focus when teaching, and whether their teaching methods work. Furthermore, it is expected for translator trainers to conduct translation assessment on regular basis in order to grade the students' performance and progress. In other words, assessing the result of students' tests and exercises is an unavoidable part in translator training, and for quite some time, the task of assessing students' performance in translation has been a confounding and largely-neglected issue (Hatim & Mason, 1997; Lauscher, 2000, Melis & Albir, 2001; Angelelli & Jacobson, 2009). Of course, the translator trainers cannot rely on their intuition and subjectivity in assessing the students' result of translation (Shiyab, 2013). Then how do the translator trainers conduct the assessment? How do they decide whether the translation is good or bad, whether the translation has conveyed the message of the source text, or whether the translation has served the purpose of the original? So far, the previous studies on this topic can be considered as scarce.

The main problem lies on the fact that there has been a lack of systematic, practical, objective, and empirical form of translation assessment (Waddington, 2001a; Kim, 2009; Williams, 2009), which leads to difficulty for translator trainers to assess the students' performance. A review of related literature revealed several models in translation assessment. Most of the models, however, only focus on assessing published translation and professional translators.

Few studies talk about the ideas of assessing translation in the area of translation pedagogy. Sainz (1994), for example, thought that the students should be involved in the process of translation assessment in order to promote students' independence and autonomy. The measurement tool proposed by Sainz (1994) is a correction chart which is divided into four columns with four labels: Mistakes, Possible Correction, Source, and Type of Mistake. Other experts such as Waddington (2001b), Khanmohammad and Osanloo (2009), and Angelelli (2009) believed that scoring rubric is the most appropriate instrument to assess the performance of translation students, and the rubrics they proposed are justified by existing linguistics and translation theories. Bowker (2000; 2001; 2003) and Baer and Bystrova-McIntyre (2009), on the other hand, perceived corpus as the most appropriate tool in studying and assessing translation works produced by the students. The models brought forward by the experts mentioned above have their own advantages and disadvantages, and they will be discussed further in Chapter 2.

However, after reviewing the literature, none of the assessment model can provide both scores and feedback for the students. Moreover, some of the assessment models only focus on surface-level linguistics and do not analyze the conveyance of

meanings. Translation assessment for pedagogical purposes ideally should be able to generate scores in order to determine whether the students deserve to pass the class or not, and this is in line with the statement by Shiyab (2013). In addition, ideally it should be able to provide necessary feedback for the students to improve their translation skills. Moreover, it should not only focus on simple linguistics errors but also errors in meanings since the main concern in conducting translation is to convey the equivalent meaning from one language to another language. Nevertheless, most of the translation models that the researcher found might not fulfill the necessary requirements mentioned above.

One translation assessment model which claims to be able to compensate all those shortcomings is meaning-oriented assessment proposed by Kim (2009). The reason why the researcher was interested in this assessment model was because some of the lecturers who teach translation subjects in the researcher's institution still use forms of assessment that focus on identifying surface-level errors (e.g. grammatical errors). While it is not necessarily wrong to assess that aspect, it is also important for the lecturers to pay attention to the transfer of message and meaning from the source text to the target text.

The basis of meaning-oriented assessment is the theory of Systemic Functional Linguistics by Halliday, and the method focuses on the register, metafunction, and lexicogrammar of the texts. Kim (2009) also claimed that the model enables more detailed analysis towards translation works, and it explores not only surface-level linguistics but also the deeper aspects of meaning within the register. Furthermore, Kim (2009) claimed that meaning-oriented assessment can

generate results in the form of scores and can provide feedback for the students. Since the researcher found no account on the usage of this particular model in the area of translator training in Indonesia, this thesis attempted to fill the gap by implementing Kim's model of meaning-oriented assessment in Indonesian context.

1.2. Research Questions

This research addressed the issue of translation assessment in the area of translator training. The researcher tried to study the implementation of meaning-oriented assessment in assessing students' translation works by focusing on three issues: results of assessment, inter-rater reliability, and raters' perception on the strengths and weaknesses of the assessment. This research attempted to answer several questions as listed below:

1. What are the results of the assessment by the raters?
2. How is the inter-rater reliability based on the result of the assessment?
3. How do the raters perceive the strengths and weaknesses of the meaning-oriented assessment?

1.3. Research Objectives

Through this study, the researcher hopes to achieve three objectives. The first objective was to compare the results of assessment in order to find out the similarities and differences among the assessments produced by the raters. The second objective was to find out the reliability of meaning-oriented assessment by conducting a test of inter-rater reliability through statistical analysis. The last objective was to find out the

raters' perception on the usage of meaning-oriented assessment, particularly on the strengths and weaknesses of the assessment model. The information would eventually be used to give the final remarks on meaning-oriented assessment.

1.4. Research Significance

This research is expected to give benefits to the development of translation pedagogy in Indonesia, specifically in translation assessment, through the description of a particular model of translation assessment. Furthermore, the researcher hopes that this research would be useful for translator trainers, who need to conduct translation assessment on a regular basis. This research would give information about alternative options in assessing their students' translation works.

In addition, the researcher hopes that this research would benefit students of translation, who need to improve their skills in translating. It is hoped that this research would be able to give insights about how meaning-oriented assessment works and what features of meaning-oriented assessment they need to pay attention to. Furthermore, the researcher hopes that this research would bring benefits for the institution where the researcher is working. By reading this research report, the institution and the lecturers can consider whether or not meaning-oriented assessment is appropriate to be implemented in the institution.

1.5. Scope of Research

This research focused on the implementation of meaning-oriented assessment proposed by Kim (2009), and the assessment is based on Halliday's Systemic

Functional Linguistics. This research also attempted to test the principle of reliability out of five principles of assessment proposed by Brown (2004). The subjects involved in this research were the former 7th semester students from the institution where the researcher is working. Ten students were selected from the total 23 students (the method of selection will be described in detail in Chapter 3), and those students were asked to translate a piece of editorial text entitled *Don't Let Drugs Rule the Nation* taken from *the Jakarta Post* (see Appendix 1).

1.6. Limitation of Research

This research has two limitations. First, to the best of the researcher's knowledge, there has been no study regarding the usage of meaning-oriented assessment in translation teaching in Indonesia. There are other studies conducted in Indonesian context which focus on translation assessment in translator training, yet all of them discuss about other forms of assessment. The researcher, however, did not manage to find any accounts or previous studies about the implementation of meaning-oriented assessment in Indonesia. As a result, the researcher faced difficulties in comparing the result of this study with previous research. However, the lack of previous studies is considered common in groundbreaking studies, and the researcher hopes that the results generated from this research can be accepted as preliminary findings which require further and more careful investigation from future researchers.

Second, the number of data used in this study is limited due to the complexity of the assessment and the participation of the rater. This may result in two further

problems. First, since the researcher intended to determine the inter-rater reliability through statistical means, the result might not be statistically significant and might not be able to provide strong evidence to establish a claim due to the limited number of data. In order to remedy this limitation, the researcher employed non-parametric test so that the statistical result can be gained despite the small number of data. Second, since the data were limited, the result of this study cannot be generalized towards all cases of translation assessment. This non-representativeness of the result, unfortunately, is a feature that needs to be accepted by the readers of this thesis.

1.7. Organization of Writing

As noted, the first chapter of this thesis discusses the background of this study, research questions, research objectives, research significances, scope and limitation of the research, and organization of the writing. The second chapter consists of discussions about varieties of translation assessment in translator training, Systemic Functional Linguistics and meaning-oriented assessment, criteria of assessment tools, some related previous studies, and theoretical framework of the research. The third chapter consists of detailed description of the methodology used in this research, starting from the research paradigm, research design, validity issues and data sources, procedures for data collection and data analysis. The fourth chapter is comprised of data findings and data analysis, and the fifth chapter consists of conclusion of the research and suggestions for future research.

CHAPTER 2

LITERATURE REVIEW

While previous studies about translation assessment in pedagogical area have discussed different types of methodology of translation assessment, it is remained unknown whether meaning-oriented assessment would be beneficial in terms of assessing students' translation work in Indonesian context. In this chapter, the researcher explains several literatures related to this study. This chapter consists of several sub-chapters or sections. Section 2.1 is about translation assessment in translator training; section 2.2 talks about Systemic Functional Linguistics, which is used as the basis for meaning-oriented assessment; section 2.3 discusses about meaning-oriented assessment and its justification and implementation; section 2.4 explains about criteria of assessment tools; section 2.5 reviews some previous studies related to translation assessment in Indonesian context; and section 2.6 consists of theoretical framework.

2.1. Translation Assessment in Translator Training

Translation assessment can be defined as a means in evaluating and determining the value of a translation, and it can be conducted for various purposes (Melis & Albir, 2001; Williams, 2009; Angelelli & Jacobson, 2009) as explained in Chapter 1. In the area of translator training, translation assessment is an unavoidable part that needs to be conducted by translator trainers. To accommodate the need, some scholars proposed a range of models and frameworks with different underlying theories.

Recent studies suggest several methods in assessing translation works. Bowker (2000; 2001; 2003), Pearson (2003), and Baer and Bystrova-McIntyre (2009), for instance, proposed the usage of corpus in assessing students' translation work and giving feedback to students. The rationale behind the notion is the fact that corpus is comprised of large number of natural data (Baer & Bystrova-McIntyre, 2009) and the fact that electronic corpora enables faster information retrieval compared to dictionary or parallel texts (Bowker, 2000).

One of the most prominent models involving corpus in translation assessment is the model suggested by Bowker (2001). She proposed the creation of Evaluation Corpus for assessment in translation pedagogy. Evaluation Corpus consists of four smaller corpora, i.e. the Comparable Source Corpus, the Quality Corpus, the Quantity Corpus, and the Inappropriate Corpus. Each of the corpora contains texts which are specific to one function. These corpora can be used by translator trainers to detect the "abnormalities", to pinpoint the erroneous parts in the students' translation, and to suggest better translation. Baer and Bystrova-McIntyre (2009), in line with Bowker (2001), asserted that Evaluation Corpus "may be considered a potentially promising tool for inclusion into translation training curricula" (p. 162). This means that besides assessing students' translation work, the corpus can be used to promote the practice of translator training.

However, the methodology has several shortcomings. First, despite the fact that the methodology can help the teachers to generate feedback for the students' translation task, it does not specify how the students' translation work should be graded (Bowker, 2001). Shiyab (2013) stated that it is essential for translator trainers

to quantify students' translation performance into scores in the area of translation pedagogy. However, Bowker's model cannot generate grades; it can only assist translator trainers in forming constructive feedback for the students. The second shortcoming is that the Evaluation Corpus is too time-consuming to establish since the process of compiling four corpora under the Evaluation Corpus is not an easy task. The process of selecting suitable texts and integrate them into each of the corpus will take such a long time (Baer & Bystrova-McIntyre, 2009; Colina, 2011). Lastly, the scope of the corpus is very limited. This is due to Bowker's (2001) emphasis that the Evaluation Corpus should be specific to only one *skopos* or function.

On the other hand, Waddington (2001a; 2001b), Khanmohammad and Osanloo (2009), and Angelelli (2009) emphasized on the usage of scoring rubrics designed with several specific criteria in assessing the translation works. Rubric is a type of assessment tools which are frequently used in education setting. Scoring rubrics can be described as schemes which contain various criteria and specific scales, and they are created to ensure consistency of educators or teachers in assessing students' work (Moskal, 2000).

Among the rubrics created to assess the translation works, the design proposed by Khanmohammad and Osanloo (2009) is the most unique, in the researcher's opinion. Khanmohammad and Osanloo (2009) designed the criteria of their rubrics by inquiring information from translator trainers. They administered questionnaires in order to elicit information about criteria that should be included into the rubrics. Afterwards, they identified the general trend across the responses from translator trainers and calculated the degree of importance of each criterion.

However, the scoring rubric designed by Khanmohammad and Osanloo (2009) has its own shortcomings. The first one is that there is no rationale that supports the design of the rubric. They did not give any justification on why they thought perspectives from translator trainers (instead of existing translation theories) should be given the main priority in the process of designing the criteria of the rubric. The lack of justification and reasoning, in the researcher's opinion, may give the scoring rubric a shaky ground in the process of implementation. The second shortcoming is that the rubric cannot be used to generate feedback. In translation pedagogy, it is important to let the students know which parts of their work that is erroneous and how they should deal with the errors. The rubric, however, only gives scores and provides grading. The third shortcoming is that the rubric has not been tested for its reliability and validity. According to Jonsson and Svingby (2007), it is very important to prove the reliability and the validity of an assessment tool or test instrument so that it is perceived as trust-worthy and unbiased. Khanmohammad and Osanloo (2009), however, failed to prove that their rubric is reliable and valid since they did not conduct any reliability or validity tests after designing their rubric.

In brief, both usages of corpus and scoring rubric are quite popular models in the area of translation assessment in pedagogy setting. However, as explained above, corpus cannot generate grades for students, are too time-consuming to construct, and has limited scope. Similarly, the usage of scoring rubric in translation assessment cannot generate feedbacks for the students, and often, the rubric is not tested for its reliability and validity. Due to these shortcomings, the researcher decided to study other types of translation assessment instead, and after reviewing relevant literature,

the researcher chose to use meaning-oriented assessment, which is based on the theories of Systemic Functional Linguistics.

2.2. Systemic Functional Linguistics

Systemic Functional Linguistics is a theory proposed by Halliday that centers on the idea that language is a functional unit. It treats language as a product of social construct by taking the situational and cultural contexts into account and emphasizes on the language function (the way the language is used) instead of language accuracy (Kim, 2009; Martin, 2010; Halliday & Matthiessen, 2014). Halliday and Matthiessen (2014) stated that a language can be divided into five levels of stratification, i.e. phonetics, phonology, lexicogrammar, semantics, and context (see Figure 1). Each level is connected to one another, and the level with smaller scope provides ground for the level with larger scope. For instance, phonetics and phonology are important in order to make the analysis towards lexicogrammar possible. Through lexicogrammar, the analysis on semantics can be conducted. Lastly, semantics provides information to determine the context.

The aspects of context, semantics, and lexicogrammar are interrelated with each other (see Figure 2). Context, according to Halliday & Matthiessen (2014), provides information about the situational and cultural aspects of the text, and it consists of Field, Tenor, and Mode, which are usually referred to as ‘register.’ Field identifies the topic and the nature of the text, Tenor deals with the social status and roles of the participants, and Mode concerns about the channel of communication. Each aspect of register is connected to each aspect of semantics.

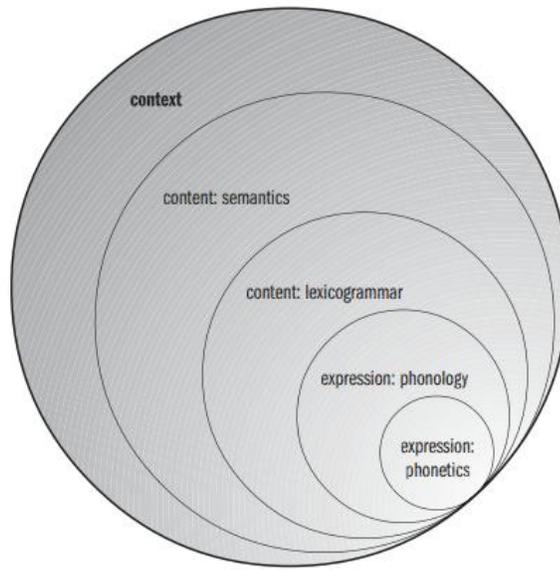


Figure 1. Stratification of language (taken from Halliday & Matthiessen, 2014, p. 26)

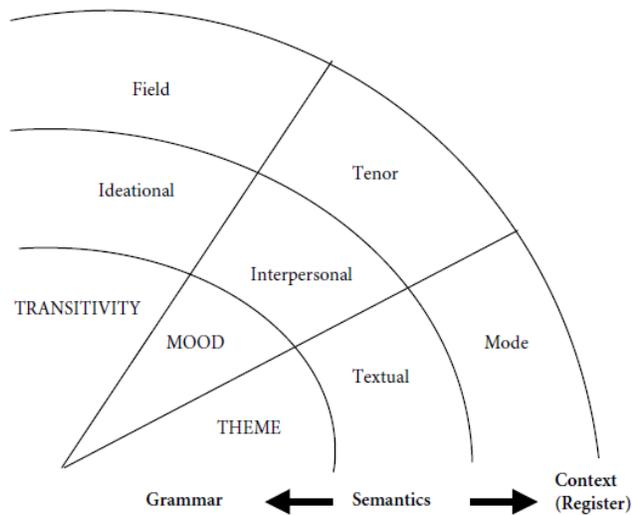


Figure 2. The relationship between grammar (lexicogrammar), semantics (metafunctions), and context (register) (taken from Kim, 2009, p. 131)

Semantics or ‘metafunctions’ is divided into three, i.e. ideational (representation of field), interpersonal (representation of tenor), and textual (representation of mode). Ideational metafunction describes humans’ experience in the real world through the usage of language, and it can be divided further into two components, i.e. experiential

(which represents experiences) and logical (which shows relationship). Interpersonal metafunction describes social and personal relationship between the speaker and the listener. Lastly, textual metafunction concerns with the cohesion and coherence of a text (Manfredi, 2011; Halliday & Matthiessen, 2014). Similar to register, each aspect of metafunctions is related to each aspect of lexicogrammar.

Lexicogrammar or ‘grammar’ consists of transitivity (realization of ideational metafunction), mood (realization of interpersonal metafunction), and theme (realization of textual metafunction). The system of transitivity includes Process (the event which is being described), Participants (people involved in the event), and Circumstances (additional information about where, when, why and how the event is happening). Meanwhile, the system of mood is reflected through modality, which “refers to the area of meaning that lies between yes and no—the intermediate ground between positive and negative polarity” (Halliday & Matthiessen, 2014, p. 691). It can be divided into two, i.e. modalization (which includes modals of probability and usuality) and modulation (which includes modals of obligation and inclination) as described in Figure 3. Lastly, theme deals with old information (theme), new information (rheme), and cohesive devices in the text (Halliday & Matthiessen, 2014).

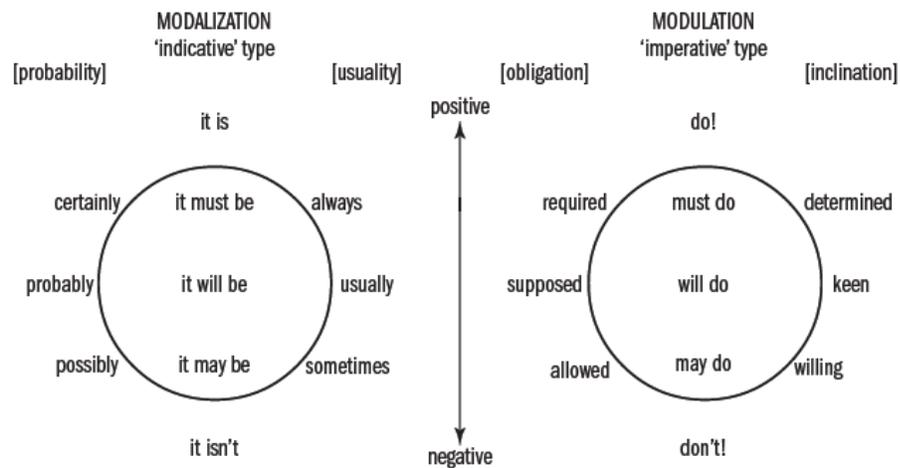


Figure 3. Modalization and Modulation (taken from Halliday & Matthiessen, 2014, p. 691)

2.3. Meaning-Oriented Assessment

Meaning-oriented assessment is a model of translation assessment for pedagogical purpose designed by Kim (2009) with Halliday's Systemic Functional Linguistics (SFL) as the underlying theory. The justification behind the usage of SFL theory because SFL pays considerable amount of attention on the meaning of a text, and this is in line with the practice of translation assessment which emphasizes on successful transfer of meaning (Newmark, 1987; Halliday, 1992; Manfredi, 2011). The criteria for meaning-oriented assessment are acquired from Halliday's SFL (Halliday & Matthiessen, 2014). Kim (2007; 2009; 2010) adopted the theory of register and three metafunctions from Halliday & Matthiessen (2014), and to analyze those metafunctions, she utilized the elements of lexicogrammar in SFL. The relation between register, metafunction, and lexicogrammar can be seen in the figure below.

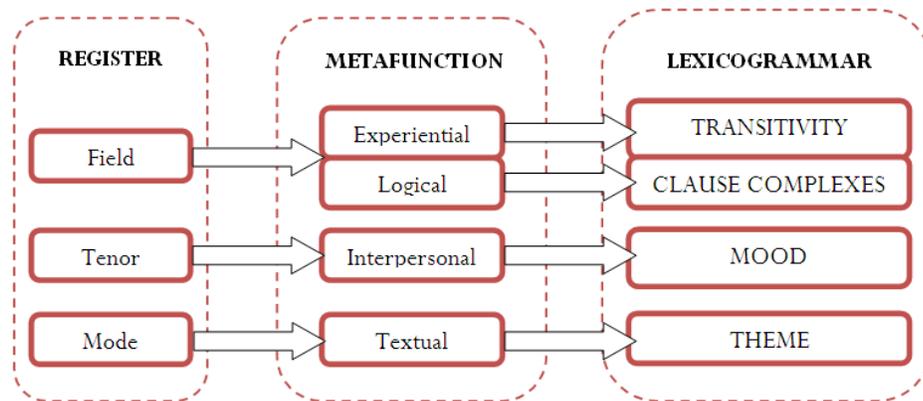


Figure 4. Relation among Register, Metafunction, and Lexicogrammatical Elements (adapted from Kim (2009))

2.3.1 Rationale and Implementation of Meaning-Oriented Assessment

Meaning-oriented assessment is created to complement Australian professional assessment by NAATI (The National Accreditation Authority for Translators and Interpreters). Kim (2009) argued the system of point reduction by NAATI only focuses on the experiential metafunction while neglecting other metafunctions. Therefore, she designed a complementary model of translation assessment based on SFL. Based on these justifications, Kim (2009) claimed that meaning-oriented assessment can generate both scores and feedback for the students.

There are two parts in the model of meaning-oriented assessment. The first part is the identification and analysis of errors, whether the errors are major or minor. Major errors involve errors in transferring experiential, ideational, interpersonal, and textual metafunctions from the source text into the target text. Those errors are then assessed for their severity in terms of accuracy and naturalness. Larson (1984) clarified that a translation is accurate when the meaning from the source text is exactly conveyed without adding, omitting, or changing any sense in the meaning. On

the other hand, a translation can be considered as natural if the translation uses grammatical forms or construction and word pairs or collocation that are adequate and acceptable in target language. After deciding whether the error is a case of inaccuracy or unnaturalness, the error needs to be analyzed more closely in order to detect the impact of the error. As seen in Figure 5, the errors might impact the level of lexis (words), clause, or text. If the error is limited to lexical level and does not affect the clause or text, the range of point reduction below lexis column should be used. If the error influences the whole clause but not the whole text, the range of point reduction below clause column should be utilized. If the error affects the whole text, the range of point reduction below text column should be used. On the other hand, the minor errors are mistakes in graphology or typography and mistakes in grammar which do not affect the transfer of experiential, logical, interpersonal, and textual meanings (see Figure 5).

| | | | Lexis | Clause | Text |
|-------|---|-------------|---------|---------|---------|
| Major | Experiential | Accuracy | 1-2 pts | 2-3 pts | |
| | | Naturalness | 1-2 pts | 2-3 pts | |
| | Logical | Accuracy | | 1-3 pts | |
| | | Naturalness | | 1-3 pts | |
| | Interpersonal | Accuracy | 1-2 pts | | 3-5 pts |
| | | Naturalness | 1-2 pts | | 3-5 pts |
| | Textual | Accuracy | | 1-2 pts | 3-5 pts |
| | | Naturalness | | 1-2 pts | 3-5 pts |
| Minor | Graphological mistakes such as spelling | | | 0.5 | |
| | Minor grammar mistakes that do not impact meaning | | | 0.5 | |

Figure 5. Criteria for Meaning-oriented Assessment (taken from Kim (2009, p. 136))

The second part is the process of generating the scores. Kim adopted the system of point reduction from NAATI professional assessment (see Figure 5), which

reduces scores from the total of 45 points based on the severity of the errors. For example, if a student inaccurately transfers the experiential meaning and that error only affects the level of lexis, the translation trainers can deduct 1 to 2 points out of 45 points according to the severity of the error.

There are several limitations of meaning-oriented assessment that need to be acknowledged. First, Kim (2009) stated that meaning-oriented assessment has not been tested for the reliability and validity. However, to compensate this shortcoming, Kim has conducted the test of pedagogical efficacy on the meaning-oriented assessment by using survey, NAATI recommendation ratio, and students' learning journals. In other words, even though this model has not been tested for its reliability and validity, its effectiveness in pedagogical context has been proven. Nevertheless, the gap of reliability and validity cannot be ignored. Therefore, this study attempted to fill the gap of Kim's study by conducting a test to find out the reliability of the assessment model.

The second shortcoming is that meaning-oriented assessment is difficult to implement if the language pairs are not described in the terms of SFL. There have been attempts by Caffarel, Martin, and Matthiessen (2004) to describe the functional profiles of languages other than English such as German, Japanese, Chinese, Tagalog, and so on. In Indonesian context, unfortunately, the researcher had not managed to find any research which describes the functional profiles of Indonesian. However, Kim (2009) managed to conduct her analysis in English-Korean language pairs even without the functional profiles of Korean at that time. The researcher, therefore,

concluded that this research is possible to conduct even with the absence of SFL-based description of Indonesian.

The third shortcoming is that the criteria proposed by Kim do not contain any descriptors. This may cause difficulties for translator trainers who will use the criteria in assessing students' translation works. Therefore, in order to mitigate this shortcoming, the researcher composed a guide for meaning-oriented assessment so that the translator trainers would not experience any confusion or difficulties in using the criteria of assessment. The guide consists of the explanation, description, and examples of each item in the criteria and the implementation of meaning-oriented assessment.

2.3.2. Example of Error Analysis in Meaning-Oriented Assessment

In meaning-oriented assessment, the identification and analysis of errors are conducted based on SFL. Each clause can be analyzed for its four strands of meanings or metafunctions (experiential, logical, interpersonal, and textual). As a result, there is a possibility that a clause may have more than one metafunctional error. Have a look at the chunk of text and its translation below.

SL : The Egyptian government **must** ensure that security forces cooperate fully and transparently (*New York Times*).

TL : Pemerintah Mesir **seharusnya** memastikan bahwa pasukan keamanan bekerja sama secara penuh dan transparan (translation from the researcher).

The aspect that needs further analysis in the translation above is the translation of modal 'must' in English into adverbial 'seharusnya' in Indonesian. Alwi (1992) stated that modals in English are frequently expressed in the form of verbs or

adverbials in Indonesian. Therefore, the difference in terms of parts of speech is not the issue in this case since English and Indonesian have different grammatical systems.

The main problem is the transfer of meaning in the translation of modal ‘must’ in English into adverbial ‘seharusnya’ in Indonesian. According to Kim (2007, 2009, 2010), the usage of modals is included into the system of mood, which affects the interpersonal metafunction. Looking at the chart of Modalization and Modulation (see Figure 3), it can be seen that modal ‘must’ followed with verb is considered as modal of obligation (in the area of Modulation) at the top position, which indicates positive or highest degree. This is due to the fact that ‘must’ has the highest level of obligation, and its usage leaves no room for arguments (Quirk, Greenbaum, Leech & Svartvik, 1985).

However, the modal ‘must’ is translated into ‘seharusnya’. According to Jayanti (2012), “seharusnya” is located in the middle position, which indicated less positive or median degree of obligation. This mismatch of position in the SL and TL as seen from Figure 3 shows that there is an inaccuracy in the translation of modal ‘must’ from SL into TL, which results in compromise of interpersonal metafunction. This inaccuracy of translation is considered severe since the translation deviates from the SL. Therefore, based on the criteria of assessment (Figure 5), 2 points (or the maximum score) in the aspect of interpersonal metafunction reflected through lexis must be deducted from the total 45 points.

2.4. Criteria of Assessment Tools

Harris (1969) mentioned three criteria that an assessment has to fulfill, i.e. validity, reliability, and practicality whereas Brown (2004) stated that there are five, i.e. practicality, reliability, validity, authenticity, and washback. An assessment is considered practical if it is cost-effective, time-effective, easy to administer, and specifically scored; reliable if it is able to generate consistency of result across occasions, students, and raters; valid if it actually tests the subject matters that are supposed to be measured; and authentic if it reflects the real world (Brown, 2004). Lastly, washback concerns about the effect of the usage of assessment towards the participants (Harris, 1969; Brown, 2004).

In this study, the researcher would only focus on the criterion of reliability since reliability is one of the elements which Kim (2009) has not explored. Kubiszyn and Borich (2013) explained that an assessment can be considered as reliable if it produces consistent result when it is administered repeatedly on different days to people with the similar proficiency level. It is important for an assessment tool to be tested for its reliability so that the users can wholly depend on the tool to give assessment for various purposes. Bachman (1990) also added that when the condition of reliability is fulfilled, the condition of validity can be achieved as well since it is believed that these two aspects complement each other.

Reliability can be divided into several types. One of them is inter-rater reliability, which can be defined as the degree of consistency of the scores generated by different raters (Bachman, 1990; Brown, 2004). If more than one raters use the same assessment model or tool in assessing one student's performance, and the result

of the raters is consistent with one another, the assessment model or tool can be considered as reliable.

In determining the reliability of an assessment, statistical calculation is often used. Saldanha and O'Brien (2014) suggested using Cohen's kappa coefficient when there are two variables and Fleiss' kappa coefficient when there are more than two. In addition, Corder and Foremen (2009) suggested using one-way ANOVA formula when dealing with large number of data and Kruskal-Wallis test with small number of data.

2.5. Previous Studies on Translation Assessment in Indonesian Context

Few studies have been conducted in Indonesia in regards to translation assessment in the area of translator training. One study that the researcher considered useful is the research by Taufik (2014). This study tested the reliability of the marking scheme that he had created based on House's Functional-Pragmatic Model of Assessment. To determine the reliability, Taufik (2014) identified the raters' perception of the scheme by using questionnaires, and the result was calculated using Kruskal-Wallis test to test the equality of perception across the raters. The calculation shows the raters have equal perception of the scheme, and the scheme is reliable.

Taufik's (2014) research is useful for this study because the study attempted to test the reliability of certain translation assessment models. Similar to Taufik, the researcher utilized statistical calculation in order to determine the reliability of meaning-oriented assessment. Therefore, the researcher included Taufik's research paper as previous study for this research.

2.6. Theoretical Framework

The figure below presents the theoretical framework of this study.

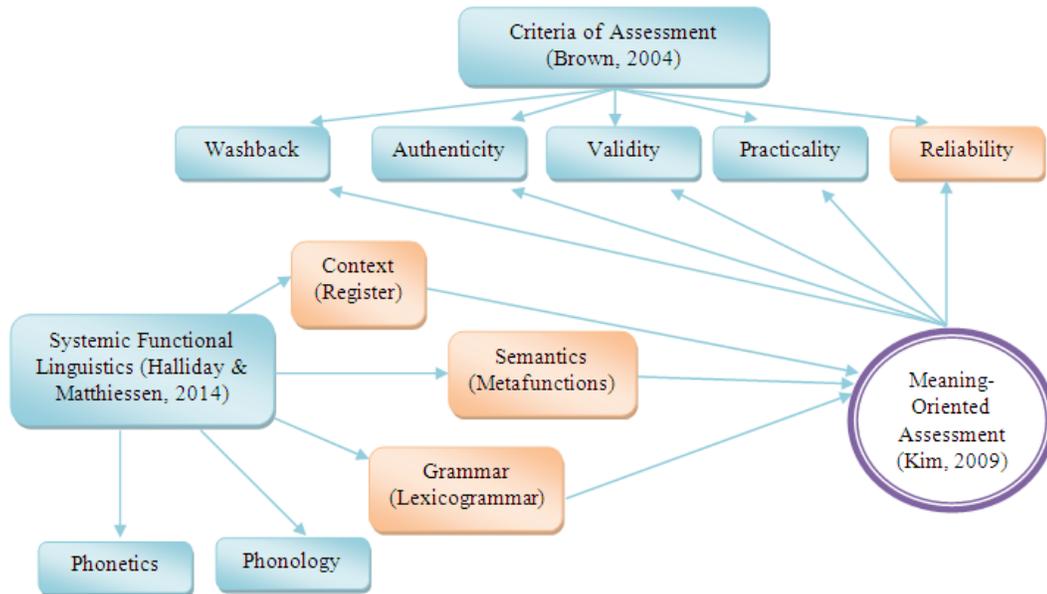


Figure 6. Theoretical Framework

Meaning-oriented assessment by Kim (2009) adapted some features of Systemic Functional Linguistics (Halliday & Matthiessen, 2014), i.e. the concepts of context (register), semantics (metafunctions), and grammar (lexicogrammar). Meanwhile, as an assessment tool, meaning-oriented assessment has to be tested based on several criteria suggested by Brown (2004). In this study, however, the researcher only focused on determining the reliability of meaning-oriented assessment model in addition to the raters' perception about the strengths and weaknesses of the assessment model.

CHAPTER 3

RESEARCH METHODOLOGY

In this chapter, the researcher explained the research methodology used to conduct this study. Complementing the theoretical review discussed in the previous chapter, the researcher presented the empirical consideration in more detailed manner. Furthermore, the researcher provided arguments and justification on why certain methods were chosen to be used in this study.

3.1 Research Paradigm

The researcher employed mixed method as the research paradigm. Mixed method is a research paradigm that emphasizes on the usage of qualitative and quantitative approaches, instruments, or data in one single research (Johnson, Onwuegbuzie & Turner, 2007; Dörnyei, 2007; Heyvaert, Maes & Onghena, 2011; Creswell, 2013). Some researchers tend to favor mixed method as it improves the validity of the research and enhances the strengths while compensating the weaknesses of the research (Dörnyei, 2007).

In this study, the researcher needed to answer three research questions, and each question needed different approach in getting the answer. The first research question is about three raters' results in assessing students' translation works using meaning-oriented assessment. The answer was explained in the form of description of the assessment, so qualitative approach was used. The second research question is about inter-rater reliability which was answered through quantitative means by using statistical calculation. The last research question is about the strengths and

weaknesses of the model based on the raters' opinion, and it was answered by interviewing the raters, which is a qualitative approach. Since this study employed both quantitative and qualitative elements, the appropriate research paradigm would be the mixed method.

The data were joined together in the analysis. The results of assessment from all three raters (from first research question) would provide ground in determining the value of inter-reliability (from second research question), and the value would be supported further by the result of the interview (from third research question). Then, all of them would be used to draw final conclusion on the implementation of meaning-oriented assessment.

3.2. Research Design

This research used triangulation as the research design. Denzin (1978, as cited in Johnson, Onwuegbuzie & Turner, 2007, pp. 114-115) defined triangulation as “the combination of methodologies in the study of the same phenomenon,” and it is claimed that “the bias inherent in any particular data source, investigators, and particularly method will be canceled out when used in conjunction with other data sources, investigators, and methods.” In other words, the shortcomings of one method are balanced by the advantages of other method.

The researcher chose triangulation because the researcher utilized different methods in conducting this study, i.e. description, statistical calculation, and interview. The description of the results of assessment would highlight the similarities and differences of the raters' assessment; the statistical calculation would

determine the value of inter-rater reliability; and the result of the interview would be used to determine the strengths and weaknesses of meaning-oriented assessment according to the views of the raters.

3.3. Validity Issues

In this section, the researcher would like to clarify the bias of this study, i.e. the participants of this study were the researcher's students. Due to this, subjective interference was highly probable in the research process, and it might create threat towards the validity of the research. Therefore, in order to minimize the effect of the bias and to increase the objectivity and validity of the research, the researcher did not participate in the process of assessment; instead, three raters conducted the assessment process and produced the result of assessment.

3.4. Data Sources

The primary source of data can be categorized into two. The first one is the assessment sheets that the raters had to fill after assessing the students' translation works. These assessment sheets were used to answer the first and the second research question. The second set of data was acquired through interviews with each rater in order to elicit their perceptions of the strengths and the weaknesses of the assessment model. Interview was chosen because according to Dörnyei (2007), this method is suitable to accommodate data elicitation from small number of participants. Moreover, interview would allow the researcher to probe and get more elaborated information regarding the raters' perception (Dörnyei, 2007). In this study, the

researcher employed semi-structured interview, so the researcher can both control the flow of the interview and prod for more details (Leech, 2002).

3.4.1. Selection of Research Participants

In this research, the researcher needed translated texts produced by the students for the raters to assess. The researcher focused on the former 7th semester students (batch 2011/2012). This particular batch was divided into two further majors during their course of study, i.e. teaching major and translation major. The researcher focused on the 7th semester students who chose translation major. The students had participated in most translation classes, such as theory of translation class, audiovisual translation class, and so on, which is why the researcher considered them capable in translating well.

There are 23 former 7th semester students, and since the researcher needed participation from other raters, it might be over-whelming for the raters to assess all 23 students' translation works. Therefore, the researcher decided to select 10 out of 23 students through TOEFL-like reading test. All 23 students were asked to do the TOEFL-like reading test in order to ensure that the selected students have enough reading skills to understand the source text in English. After the test was conducted, 10 students with the highest score were selected as the main participants of the research.

3.4.2. Selection of Translation Task

All 23 students were asked to translate a newspaper article, specifically an editorial, from English to Indonesian (see Appendix 1), and the translated texts by 10

students with the highest scores in TOEFL-like reading test were assessed by three raters. Editorial from newspaper was chosen since it can be categorized as argumentative text, and according to Tirkkonen-Condit (1985) as cited in Karoly (2006) and Kim (2009), argumentative text is difficult to be translated due to the fact that it might cause comprehension problem. Taufik (2014) also argued that argumentative text generally consists of subjective interpretations of the researcher, and it requires the utmost skills of the translators to render such text.

Considering the arguments from Tirkkonen-Condit (1985) as cited in Karoly (2006), Kim (2009), and Taufik (2014), the researcher chose to use editorials from newspaper. The editorial that the researcher chose was taken from *the Jakarta Post* newspaper since *the Jakarta Post* is considered as one of the leading English newspapers in Indonesia. The number of words in this editorial is 503 words, which is more or less similar to the number of words in Kim's model text (2009).

Before administering the translation task on the students, the researcher conducted a test on the text in order to determine the degree of readability. One of the procedures widely used in determining the difficulty of a text is Flesch-Kincaid test, particularly Flesch Reading Ease test (Kincaid, Fishburne Jr., Rogers & Chissom, 1975). Flesch Reading Ease test considers the total number of syllables, words, and sentences in the text, and the score ranges from 0 to 100. The closer the score to 0, the more difficult the text is. The researcher conducted the readability test on Flesch Reading Ease Score by utilizing the feature built in Microsoft Word 2007. The result shows that the text chosen by the researcher has the score of 42.7 for Flesch Reading Ease. This means that the text is fairly difficult to read, and according to Flesch (n.d.),

the text is suitable for college students. Therefore, the researcher considered this text as ideal for the research project.

After all 23 students translated the text, 10 students with the highest score in TOEFL-like reading test were selected. The results of the students' translation were typed and organized side-by-side with the source text in the form of assessment sheets (see Appendix 6). The assessment sheets were later distributed to the raters in digital forms.

3.4.3. Selection of Raters

For this study, the researcher would need three raters, and those raters were tasked in assessing students' translation works using meaning-oriented assessment. In order to recruit and select the raters, it is important to set some requirements so that the criteria of the selection can be retained (Gwet, 2014).

The researcher set three criteria for the selection of the raters. First, the raters have to be Indonesian citizens. This is to ensure that the raters have the capabilities to understand the grammaticality and technicalities of Indonesian. Second, the raters have been professionally paid to conduct translation works, particularly in English and Indonesian language pairs. This is to ensure that the raters are professional translators and have the capabilities to understand the grammaticality and technicalities of both English and Indonesian. Moreover, through their experience in translating professionally, their sensitivity towards language use has developed, and they already have clear mental image on how a good translation should look or sound like. Third and last, the raters have to have experiences in teaching translation

subjects in university level for at least one semester. This is important since meaning-oriented assessment is an assessment model for translation pedagogy. Through the teaching and training students in translation, it is expected that the lecturers are capable in spotting the errors in the students' translations immediately and assigning appropriate score deduction to each error that they would find in students' translation works.

3.4.4. Pilot Study

Before the researcher distributed the guide and students' translation works to the raters, the researcher conducted a pilot study in order to test the feasibility of the assessment process. The rater for this pilot study (hereafter, Rater 0) is a permanent lecturer in the institution where the researcher is working. Rater 0 is an Indonesian citizen and has been translating professionally since 1999. So far, he has taught various translation subjects such as theory of translation, audiovisual translation, literary and popular translation, legal document translation, and many more.

The researcher explained the mechanism of meaning-oriented assessment and gave the guide (which had not undergone revision) to Rater 0, and he assessed the students' translation works based on the information in the guide and the explanation the researcher gave. After the he finished assessing, the researcher conducted a short interview with Rater 0 in order to find out the shortcomings of the research design and the guide to meaning-oriented assessment.

Rater 0 suggested the researcher to organize the source text and target text side-by-side in the assessment sheets so that the form is similar to parallel text. This,

according to the rater, would simplify the rater's task and make the texts easier to view and read. Another suggestion involves giving examples to all elements of assessment. In the early version of the guide, the researcher did not put many examples in the guide, and according to Rater 0, the examples would give the raters clearer picture about the areas that the assessment covers.

Based on those suggestions, the researcher continuously revised the guide by consulting different books on SFL and journal articles on SFL in translation. Furthermore, the researcher also changed the layout of the data presentation. In the pilot study, the researcher gave the source text and target texts in the form of essay with several paragraphs to the rater. For the actual research, the researcher put the each sentence of the source text and target text side-by-side in tables as suggested by Rater 0 in the pilot study, and it can be seen in Appendix 6.

3.4.5. Guide to Meaning-Oriented Assessment

Guide to meaning-oriented assessment was a description of meaning-oriented assessment, and it contains information on the implementation of meaning-oriented assessment, the units of analysis, and some examples. The earlier version of the guide was very brief and without examples on the elements of the assessment. The researcher, based on suggestions acquired from the pilot study, decided to revise the whole guide and added more examples and explanation.

In the process of composing the guide, the researcher drew insights from Kim (2007; 2009; 2010), Halliday and Matthiessen (2014), Thompson (2014), Martin, Matthiessen, and Painter (2010), and Eggins (2004). The guide was designed

specifically to address the text that the researcher chose, and the example was also taken from the translation of the students that were not chosen as the main participants of this study. The researcher had revised the guide for many times in order to improve the readability of the text after the pilot study revealed that the guide was not comprehensive enough. The meaning-oriented assessment guide can be seen in Appendix 7.

After the guide was completed, the researcher contacted the raters to set up date and place to meet. The purpose of the meeting was to explain the mechanisms of meaning-oriented assessment. The researcher decided to individually meet the raters instead of gathering the raters together because of two reasons. First, it was quite difficult to find specific times in which all raters could gather together since each rater had his or her own schedules and works. Second, the researchers would like to prevent the raters to discuss with each other. Angelelli (2009) mentioned that when raters meet and start convening with each other, it could compromise the inter-rater reliability of the scoring because through the discussion, the raters might be influenced by other raters' views and perceptions instead of following the actual views and procedures of meaning-oriented assessment. Due to the reasons mentioned above, the researcher decided to arrange separate sessions with each rater. The information about the meetings can be seen in the Appendix 8.

3.4.6. Interview Protocol & Questions

To answer the third research questions, the researcher interviewed the raters in order to find out the strengths and weaknesses of meaning-oriented assessment. To

ensure that the researcher followed consistent procedures and did not miss any important steps in the interview, the researcher composed an interview protocol.

Jacob and Furgerson (2012, p. 2) explained:

An interview protocol is more than a list of interview questions; it also extends to the procedural level of interviewing and includes a script of what you will say before the interview, script for what you will say at the conclusion of the interview, prompts for the interviewer to collect informed consent, and prompts to remind the interviewer the information that she or he is interested in collecting.

In other words, an interview protocol is considered as a useful tool since it contains information on steps that the researcher had to take in order to complete the interview process and to ensure a fruitful interview. The interview protocol composed by the researcher consists of the title of the research, brief description of the research, the research objectives, efforts to ensure the confidentiality of the raters, several reminders for the researcher (to hand out written consent forms to the raters, to test recording device, to request the raters' permission for further contact if necessary, and so on), and the interview questions along with the probes. The interview protocol and questions can be seen in Appendix 2.

Each interview question is supplemented with probes. Probes are secondary questions which are asked in order to get more information from the interviewees, and they were only asked when there is a chance for the interviewer to explore the answers from the interviewees (Creswell, 2013).

Before conducting the interview, the researcher sent the list of interview questions without the interview protocol and the probes to the raters via email. Sending interview questions beforehand was beneficial since the raters could clarify

any difficulties in understanding the interview questions before the actual interview took place. Moreover, the raters could prepare and organize their answers so that the answers provided would be more coherent and relevant. Lastly, this would impact the length of time used for conducting the interview. Since the questions were sent prior to the interview and the raters had sufficient time to review and prepare their answers, the interview process was more time-efficient (Sutrisno, 2014).

3.5. Data Collection and Analysis

Prior to data collection, the researcher contacted the raters to ask for their willingness to participate in this study. After acquiring their permission, the researcher arranged the time to meet with all the raters for the purpose of explaining the mechanisms of meaning-oriented assessment in assessing students' translation works and giving them the guide and assessment sheets.

Afterwards, the researcher began collecting the data by asking the selected former 7th semester students to translate the editorial in Appendix 1 (translation brief is also provided) manually without any help of dictionary or other translation aids within one and half hour time limit. The result was collected and typed side-by-side with the source text. Each assessment sheet containing the table of source text and target text was coded with Student_1, Student_2, and so on.

On the day when the researcher met the raters, the researcher explained the mechanism of meaning-oriented assessment for 1 to 2 hours. Afterwards, the researchers gave the guide and sent the softcopies of the assessment sheets.

After some agreed period of time, the raters sent back the assessment sheets to the researcher. Those assessment sheets were the data used to answer the first and second research questions. By compiling, comparing, and contrasting the results of assessment, the researcher formulated answer for the first research question.

To answer the second research question, the scores from all raters for each student were tested for its normality. If the distribution of scores appears to be normal, One-way ANOVA would be used to find out the inter-rater reliability. However, if the result suggests abnormal distribution, non-parametric test, specifically the Kruskal-Wallis test, would be employed. Gwet (2014) stated that Kruskal-Wallis test can be used to determine the inter-rater reliability when the research involves three or more raters. Corder and Foreman (2009) also explained that Kruskal-Wallis test compares two or more independent samples to see whether there is difference in the data, and it is employed when the data is in ordinal scale and is less than 20.

Before conducting the calculation, the hypotheses have to be formulated. The hypotheses are:

H_0 : “There is no significant difference across the scores by the raters”

H_1 : “There is significant difference across the scores by the raters.”

If the null hypothesis is accepted, it means that the raters have equal perception and consequently, the model is reliable. On the other hand, if the null hypothesis is rejected, it means that the raters have different perceptions, and the model can be considered as unreliable. After the calculation was conducted and the hypothesis was checked, a conclusion on the reliability of the model was then drawn.

To answer the third research question, the researcher conducted semi-structured interview for approximately 30 minutes to one hour towards each rater in order to find out the strengths and the weaknesses of meaning-oriented assessment. The questions that were asked are listed in Appendix 2; however, the researcher improvised and added more questions when necessary in order to acquire more data. The whole interview were recorded and transcribed, and the transcriptions were used as the data to answer the third research question. The transcriptions were then coded and analyzed before any conclusions about the strengths and weaknesses of the model were drawn.

CHAPTER 4

FINDINGS AND DISCUSSION

This chapter will unveil the result of the research that has been described above. The chapter consists of three parts in accordance to the research questions. Section 4.1 compares and contrasts the results of assessment by all three raters. Section 4.2 summarizes the results of assessment in the form of final scores and describes the result of inter-rater reliability test. Section 4.3 explains the strengths and weaknesses of meaning-oriented assessment based on the raters' perception. Lastly, Section 4.4 expounds the result of data triangulation. Each part will be carefully explained and elaborated.

4.1. Comparison of Results of Assessment by Raters

From the overall comparison among the results of assessment by the three raters, it was found that the differences were more evident than the similarities. It was apparent from the comparison that the raters had distinctive perspectives in conducting meaning-oriented assessment. When the researcher compared the assessment by three raters, the researcher found similar cases of those distinctions. The researcher decided to group those similar cases into several classifications. The sections below are the classification on the results of assessment conducted by the raters.

4.1.1 Similar Error Categories, Similar Point Deduction

The similarities could generally be found in the categories of metafunctional errors. Some of the results by three raters show that the three raters had similar perceptions in identifying the errors in the translation, assigning the errors into their respective categories of metafunctional errors, and deciding the point deduction for the error. Below is one example of translation by Student 6 on Sentence 1.

| | |
|---|--|
| ST | Don't Let the Drugs Rule the Nation |
| TT | Jangan Biarkan Narkoba Merenggut Bangsa |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) |
| Rater 2 | Experientially Inaccurate-Lexis (2) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Unnatural-Lexis (2) Experientially Inaccurate-Clause (2) Experientially Unnatural-Clause (2) |

Table 1. Comparison of Assessment Results (Sentence 1; Student 6)

As seen from Table 1 above, all the raters agree that there was an accuracy problem in terms of experiential metafunction on the level of lexis. Rater 1, 2, and 3 gave the same point deduction for the error, which was 2 point deduction. However, it can also be seen that Rater 3 identified other types of errors as well. Similar case can also be seen from the translation by Student 5 on Sentence 2.

| | |
|---|---|
| ST | President Joko "Jokowi" Widodo's commitment to the fight against drugs is part of his responsibility to protect the nation. |
| TT | Komitmen Presiden Joko Widodo "Jokowi" untuk penentangan terhadap narkoba adalah bagian dari tanggung jawabnya dalam melindungi bangsa. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) |
| Rater 2 | Experientially Inaccurate-Lexis (2) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Unnatural-Lexis (2) |

| | |
|--|---|
| | Experimentally Inaccurate-Clause (2) Experimentally Unnatural-Lexis (2) Interpersonally Unnatural-Lexis (1) |
|--|---|

Table 2. Comparison of Assessment Results (Sentence 2; Student 5)

In Table 2, it can be seen that all the raters had identified an accuracy problem in terms of experiential metafunction on the level of lexis. All the raters suggested 2 point deduction for this error. Similar to the previous case, Rater 3 had also identified other types of errors in the translation. Below is another example of similar case.

| | |
|---|--|
| ST | He has won promotions since then. |
| TT | Sejak saat itu, dia telah memenangi banyak promosi pangkat. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experimentally Unnatural-Clause (2) |
| Rater 2 | Experimentally Unnatural-Clause (2) |
| Rater 3 | Experimentally Inaccurate-Lexis (2) Experimentally Unnatural-Lexis (2) Experimentally Inaccurate-Clause (2) Experimentally Unnatural-Clause (2) |

Table 3. Comparison of Assessment Results (Sentence 12; Student 10)

As seen in Table 3, all raters agreed that Student 10 had committed error in naturalness in terms of experiential metafunction on the level of clause. They also agreed that the error warrants for 2 point deduction. Also, similar to previous cases, Rater 3 managed to identify other types of errors in the translation. The example below also illustrates similar phenomenon.

| | |
|-----------|---|
| ST | But the way the government and law enforcers have combated drugs so far has not shifted from the textbook: Raids are conducted here and there, most of the time with much fanfare, dozens are arrested and it is business as usual again until the next cycle of raids. |
| TT | Tetapi cara pemerintah dan penegak hukum untuk melawan narkoba sejauh ini tidak ada perubahan: Pengecekan dilakukan dimana- |

| | |
|---|--|
| | mana, kebanyakan dengan adanya penggerebekkan, banyak yang ditahan dan kembali seperti biasa lagi sampai pengecekan selanjutnya. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (3) Experientially Inaccurate-Clause (3) |
| Rater 2 | Experientially Inaccurate-Lexis (2) Experientially Unnatural-Clause (2) Logically Inaccurate-Clause (3) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (2) Experientially Unnatural-Lexis (2) Experientially Unnatural-Clause (2) Textually Inaccurate-Clause (2) Textually Unnatural-Clause (2) |

Table 4. Comparison of Assessment Results (Sentence 7; Student 3)

As seen above, all raters agreed that the translation had accuracy problem in terms of experiential metafunction on the level of lexis, and they agreed that the error deserved 2 point deduction. In addition, Rater 2 and Rater 3 agreed that the translation was also problematic in the area of naturalness, particularly in terms of experiential metafunction on the level of clause, and both of them also agreed that the error deserved 2 point deduction. Other than those similarities, all raters had different opinions on other types of errors. The example below is another similar phenomenon.

| | |
|---|--|
| ST | Challenges facing drug eradication also include impunity or at least light punishments. |
| TT | Tantangan melawan eradikasi narkoba termasuk impunitas atau sedikitnya hukuman ringan. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (3) |
| Rater 2 | Experientially Inaccurate-Lexis (2) |

| | |
|----------------|---|
| | Experientially Inaccurate-Lexis (2) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (2) Experientially Unnatural-Lexis (2) Experientially Unnatural-Clause (2) |

Table 5. Comparison of Assessment Results (Sentence 10; Student 5)

Table 5 above illustrates the comparison of assessment results by the raters on the translation of Sentence 10 by Student 5. As seen in the table, all raters had identified two accuracy problems in terms of experiential metafunction on the level of lexis. Similar to the previous cases, all three raters agreed that the errors deserved 2 point deduction. However, it can also be seen that Rater 1 and 3 also identified accuracy problem in terms of experiential metafunction on the level of clause, yet they gave different point deduction. Furthermore, Rater 3 also identified naturalness problem in term of experiential metafunction on the level of lexis and clause.

4.1.2 Similar Error Categories, Different Point Deduction

Another finding in the results of assessment is the similarities in term of error categories yet differences in term of point deduction. The similarities lie on the categorization of errors by the raters. The same errors, however, were regarded differently when the raters assigned point deduction. The point deduction is assigned based on the severity of the errors, and it is believed that the raters are capable of analyzing the severity of the errors and thus, able to deliver point deduction accordingly. Nevertheless, it is evident from the comparison that the raters had

different ideas about the severity of the errors. Below is one example taken from the results of assessment on Sentence 1 by Student 5.

| | |
|---|--|
| ST | Don't Let the Drugs Rule the Nation |
| TT | Jangan Biarkan Narkoba Mengatur Bangsa |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) |
| Rater 2 | Experientially Inaccurate-Lexis (1) |
| Rater 3 | Experientially Inaccurate-Lexis (2) |
| | Experientially Unnatural-Lexis (2) |
| | Experientially Inaccurate-Clause (2) |
| | Experientially Unnatural-Clause (2) |

Table 6. Comparison of Assessment Results (Sentence 1; Student 5)

As seen in Table 6, Rater 1, 2 and 3 believed that the translation of Sentence 1 by Student 5 was problematic. They also believed that the problem was related to the inaccuracy in experiential metafunction on the level of lexis. The difference, as mentioned before, is the assignment of point deduction. Rater 1 and Rater 3 assigned 2 point deduction whereas Rater 2 assigned 1 point deduction. Other difference is that Rater 3 managed to identify other types of errors as well. Below is another example of the similar case on the translation of Sentence 1 by Student 10.

| | |
|---|--|
| ST | Don't Let the Drugs Rule the Nation |
| TT | Jangan Biarkan Narkoba Menjajah Negara ini |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) |
| Rater 2 | Experientially Inaccurate-Lexis (2) |
| Rater 3 | Experientially Inaccurate-Lexis (1) |
| | Experientially Unnatural-Lexis (1) |
| | Experientially Inaccurate-Clause (2) |
| | Experientially Unnatural-Clause (2) |

Table 7. Comparison of Assessment Results (Sentence 1; Student 10)

Similar to the previous case, it can be seen that all the raters agreed that the error was a problem of inaccuracy in the experiential metafunction on the level of lexis. However, the point deduction is again different. Rater 1 and Rater 2 believed that the error was severe, so they assigned 2 which is the maximum point deduction for the case. Rater 3, on the other hand, the error was not severe, so he gave 1 point deduction, which is the minimum point deduction. Moreover, similar to previous case, Rater 3 also managed to identify other kinds of errors in the translation by Student 10.

The cases above occurred when the assessment involved translation of one simple sentence, which is composed of only one independent clause. In the case of the translation of longer sentences or sentences which are composed of more than one clause, all raters usually listed more than one category of errors. In these cases, few similarities could be spotted. However, the differences among the raters also became more and more evident. Below is an example of the assessment on the translation conducted by Student 2 on Sentence 3.

| | |
|---|--|
| ST | Last year he repeatedly warned of “a drug emergency”, saying 50 people died in relation to drug use every day and 4.5 million people were addicted, giving justification to execute drug convicts. |
| TT | Pada tahun lalu, Beliau berulang kali menyatakan “Darurat Narkoba”, dan mengatakan bahwa 50 orang meninggal dunia setiap harinya dikarenakan narkoba dan 4,5 juta orang mengalami kecanduan narkoba. Eksekusi terhadap tersangka pengedar narkoba akan menjadi keadilan atau penanganan masalah ini. |
| Error Categories (Point Deduction) | |
| Rater 1 | Minor Error (0.5) Minor Error (0.5) Experientially Inaccurate-Lexis (2) Experientially Unnatural-Lexis (1) Experientially Inaccurate-Clause (3) |

| | |
|----------------|---|
| Rater 2 | Experientially Inaccurate-Lexis (1) Logically Inaccurate-Clause (2) Logically Unnatural-Clause (2) |
| Rater 3 | Experientially Inaccurate-Lexis (1) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (2) Logically Inaccurate-Clause (3) Logically Unnatural-Clause (3) Textually Unnatural-Text (3) |

Table 8. Comparison of Assessment Results (Sentence 3; Student 2)

It can be seen from Table 8 that there are only few similarities among the assessments by all the raters. The similarities are only limited to the types of error categories. For example, all three raters agreed that the translation is problematic in terms of experiential metafunction, specifically problems of accuracy on the level of lexis. However, the differences lie on the fact that each rater had his or her opinion in regards to the point deduction that should be given. Rater 1 gave 2 point deduction whereas Rater 2 gave 1 point deduction. Rater 3, on the other hand, found 2 occurrences of such error and gave 1 point deduction to one occurrence and 2 point deduction for the other occurrence.

Moreover, Rater 2 and 3 identified accuracy and naturalness problems in terms of logical metafunction on the level of clause. For the accuracy problem, Rater 2 gave 2 point deduction whereas Rater 3 gave 3 point deduction. Similarly, for the naturalness problem, Rater 2 gave 2 point deduction, and Rater 3 gave 3 point deduction.

Furthermore, from Table 8, it can be seen that Rater 3 particularly spotted more errors compared to the other two raters, and Rater 2 identified the least number

of error types in his assessment. Rater 1 was the only one who found minor errors in the translation, and Rater 3 was the only one who found the error in naturalness in term of textual metafunction on the level of text. Another example of translation of longer sentences with similar error categories but different point deduction can be seen in the following table.

| | |
|---|--|
| ST | At the opening of a Cabinet meeting on Wednesday, the President said he wanted “more intensive, braver, crazier and more comprehensive integrated efforts to eradicate drugs”. |
| TT | Pada pembukaan rapat Kabinet hari Rabu lalu, Presiden menginginkan usaha yang terorganisir dalam memerangi narkoba, usaha yang lebih intens, berani, gila dan mendalam. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (3) Experientially Unnatural-Lexis (1) |
| Rater 2 | Experientially Inaccurate-Clause (2) |
| Rater 3 | Experientially Inaccurate-Lexis (1) Experientially Inaccurate-Lexis (1) Experientially Inaccurate-Clause (2) Experientially Unnatural-Lexis (1) Experientially Unnatural-Clause (2) Minor Error (0.5) |

Table 9. Comparison of Assessment Results (Sentence 4; Student 10)

In Table 9, it can be seen that Rater 1 and Rater 3 agreed that the translation had 2 occurrences of accuracy problem in terms of experiential metafunction on the level of lexis. Rater 1 gave 2 point deduction for each occurrence whereas Rater 3 gave 1 point deduction for each occurrence.

It can also be seen that all three raters agreed that within the translation, there was an accuracy problem in terms of experiential metafunction on the level of clause.

However, all three raters assigned different point deduction. Rater 1 gave 3 point deduction whereas Rater 2 and Rater 3 gave 2 point deduction. Below is another example of the similar phenomenon.

| | |
|---|--|
| ST | A debate over this possibility should start sooner rather than later, simply because the war on drugs has already begun. |
| TT | Debat tentang hal ini harus segera dimulai, lumrah karena perang terhadap narkoba sudah dimulai. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (1) |
| Rater 2 | Experientially Inaccurate-Lexis (1) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (2) Experientially Unnatural-Lexis (2) Experientially Unnatural-Clause (2) Interpersonally Inaccurate-Lexis (1) Interpersonally Inaccurate-Text (3) Textually Inaccurate-Clause (2) Textually Unnatural-Lexis (2) Textually Unnatural-Text (3) |

Table 10. Comparison of Assessment Results (Sentence 20; Student 10)

As seen in Table 10 above, all raters had identified accuracy problem in terms of experiential metafunction on the level of lexis. Rater 1 and Rater 2 believed the error deserved 1 point deduction whereas Rater 3 thought the error deserved 2 point deduction. Furthermore, it can be seen that Rater 3 managed to identify more types of errors.

4.1.3 Different Error Categories, Different Point Deduction

Among the results of assessment by all three raters, the researcher also found one case in which the raters had completely different opinions altogether in terms of

error categories. Since the error categories were different, the point deduction generated by the raters was also different. Below is the case by Student 1 on the translation of Sentence 2.

| | |
|---|---|
| ST | President Joko “Jokowi” Widodo’s commitment to the fight against drugs is part of his responsibility to protect the nation. |
| TT | Komitmen Presiden Jokowi dalam memerangi narkoba adalah bagian dari tanggung jawabnya untuk melindungi negara. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) |
| Rater 2 | Minor Error (0.5) Experientially Unnatural-Lexis (1) |
| Rater 3 | Interpersonally Unnatural-Lexis (1) |

Table 11. Comparison of Assessment Results (Sentence 2; Student 1)

As seen in Table 11 above, all raters had completely different classification on the errors that they found in the translation. Rater 1 believed that there were two instances of accuracy problem in terms of experiential metafunction on the level of lexis. Rater 2 believed that there were a minor error and a naturalness problem in terms of experiential metafunction on the level of lexis. Lastly, Rater 3 believed that there was a naturalness problem in terms of interpersonal metafunction on the level of lexis.

4.1.4 No Error Categories

Another phenomenon found after comparing the assessment results by three raters was that one of the raters did not consider the translation of the sentence was problematic whereas the other raters believed the opposite. Below is an example of translation conducted by Student 1 on Sentence 1.

| | |
|---|---|
| ST | Don't Let the Drugs Rule the Nation |
| TT | Jangan Biarkan Narkoba Menguasai Negara |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) |
| Rater 2 | Experientially Inaccurate-Lexis (1) |
| Rater 3 | (No error) |

Table 12. Comparison of Assessment Results (Sentence 1; Student 1)

As seen in Table 12, Rater 1 and Rater 2 agreed that the translation was inaccurate in terms of experiential metafunction on the level of lexis. However, they had different perception on the severity of the error. Rater 1 gave 2 point deduction whereas Rater 2 gave only 1 point deduction. Rater 3, however, believed that the translation was error-free and acceptable. A similar case can be seen in the translation of Sentence 2 by Student 3.

| | |
|---|---|
| ST | President Joko "Jokowi" Widodo's commitment to the fight against drugs is part of his responsibility to protect the nation. |
| TT | Presiden Joko "Jokowi" Widodo berkomitmen untuk melawan narkoba sebagai tanggungjawabnya untuk melindungi negara. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) |
| Rater 2 | (No error) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (2) Interpersonally Unnatural-Lexis (1) |

Table 13. Comparison of Assessment Results (Sentence 2; Student 3)

As seen in Table 13, Rater 1 believed that the sentence was problematic in terms of experiential metafunction, specifically inaccuracy in the level of lexis, and she thought the error deserved 2 point deduction. Rater 3 also had similar belief as Rater 1 regarding the classification of error and assignment of point deduction. However, Rater 3 believed there were other errors in the translation. In his opinion, Student 3

also committed error of accuracy in experiential metafunction on the level of clause and error of naturalness in interpersonal metafunction on the level of lexis. In total, Rater 3 suggested 5 point deduction. Rater 2, however, believed that Student 3 had rendered Sentence 2 well, and no errors could be found in the translation. Another similar case can be found in the translation of Sentence 2 by Student 8.

| | |
|---|--|
| ST | President Joko “Jokowi” Widodo’s commitment to the fight against drugs is part of his responsibility to protect the nation. |
| TT | Presiden Jokowi Widodo berkomitmen untuk perang melawan narkoba karena ini merupakan kewajiban seorang kepala negara dalam melindungi bangsa. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (1) |
| Rater 2 | (No error) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Unnatural-Lexis (2) Experientially Unnatural-Lexis (2) Experientially Inaccurate-Clause (2) Experientially Unnatural-Clause (2) Logically Inaccurate-Clause (2) Logically Unnatural-Clause (2) Interpersonally Unnatural-Lexis (1) |

Table 14. Comparison of Assessment Results (Sentence 2; Student 8)

In this case, differences among the raters are very apparent. Rater 1 and Rater 3 believed that there was an accuracy problem in terms of experiential metafunction on the level of lexis. However, despite the same category, both gave different point deduction. Rater 1 gave 1 point deduction whereas Rater 2 gave 2 point deduction. Furthermore, it could be seen in Table 14 that Rater 3 identified other types of errors whereas Rater 2 did not find the translation problematic. Below is another similar case on translation of Sentence 12 by Student 1.

| | |
|---|---|
| ST | He has won promotions since then. |
| TT | Ia bahkan dipromosikan sejak saat itu. |
| Error Categories (Point Deduction) | |
| Rater 1 | (No error) |
| Rater 2 | (No error) |
| Rater 3 | Experientially Inaccurate-Lexis (1) Experientially Inaccurate-Clause (2) |

Table 15. Comparison of Assessment Results (Sentence 12; Student 1)

As seen in Table 15, Rater 1 and Rater 2 did not find any errors in the translation of Sentence 12. However, Rater 3 believed there were accuracy problems in terms of experiential metafunction on the level of lexis and clause. The example below illustrates similar phenomenon.

| | |
|---|--|
| ST | Many would remember an Army officer, a general's son, who was found guilty of drug abuse more than 15 years ago but had his prison sentence reduced and his dismissal from the military revoked. |
| TT | Banyak yang akan mengingat seorang petugas tentara, anak jenderal, yang bersalah dalam penggunaan narkoba lebih dari 15 tahun yang lalu tetapi mendapat pengurangan masa hukuman penjara dan dikembalikan ke satuan militer. |
| Error Categories (Point Deduction) | |
| Rater 1 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (3) Experientially Unnatural-Lexis (1) Experientially Unnatural-Lexis (1) Experientially Unnatural-Lexis (2) |
| Rater 2 | (No error) |
| Rater 3 | Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Lexis (2) Experientially Inaccurate-Clause (2) Experientially Unnatural-Lexis (2) Experientially Unnatural-Clause (2) Logically Inaccurate-Clause (2) Logically Unnatural-Clause (1) |

Table 16. Comparison of Assessment Results (Sentence 11; Student 9)

Based on Table 16, it is evident that in Rater 2's opinion, there were no errors found in the result of translation. However, Rater 1 and Rater 3 believed differently. Rater 1 believed that there were five occurrences of errors, and those errors deserved 9 point deduction in total. On the other hand, Rater 3 thought there were more errors in the translation. According to Rater 3, there were seven occurrences of errors in the translation of Sentence 11 by Student 9, and Rater 3 believed those errors deserved 13 point deduction in total.

Based on the findings for the first research question, it is clear that there is a great deal of differences and disagreement among the raters compared to the amount of similarities. There are two possible reasons for this occurrence.

The first possible reason is that the raters possess almost no prior knowledge of Systemic Functional Linguistics (SFL). Meaning-oriented assessment focuses heavily on the aspects of SFL, and complete understanding of the concepts of SFL would assist the raters in conducting the assessment. It is important that the raters could differentiate among experiential, logical, interpersonal, and textual metafunctions and could recognize elements associated with those metafunctions. Experiential metafunction, for example, is reflected by the element of transitivity; logical metafunction is evident through the relationship between clauses/clause complexes; interpersonal metafunction is related to the mood of the text; and textual metafunction is apparent through the theme of the text. Understanding on these metafunctions and components related to them could greatly boost the effectiveness of meaning-oriented assessment.

However, as mentioned previously, the raters had little to no prior knowledge on SFL, and since the concept of meaning-oriented assessment is built on the SFL theories, there is a possibility that the raters might have experienced problems in applying meaning-oriented assessment due to lack of knowledge on SFL. The problems might involve the difficulties in determining the appropriate categories of errors (whether the errors belong to experiential, logical, interpersonal, or textual metafunction), the nature of errors (whether the errors are problems of accuracy or naturalness), the level of errors (whether the errors affect the level of lexis, clause, or text), or the severity of the errors (which determines the amount of point deduction). The ranges and types of difficulties faced by the raters are extensive, and it would greatly expand the previous studies if those problems could be identified and classified. Based on that understanding, the researcher included questions which could elicit the raters' perception on meaning-oriented assessment for the interview sessions with the raters. The results of the interviews would be discussed further in Section 4.3.

The second possible reason is related to the first possible reason. The raters might not receive enough information on SFL from one single training session. As mentioned in Chapter 3, the researcher only met the raters once to discuss the mechanisms and procedures of meaning-oriented assessment. During that one- to two-hour training, there is a probability that the raters did not fully gain complete understanding on the concepts of SFL and meaning-oriented assessment. Coupled with the fact that the raters had little to no prior knowledge on SFL and meaning-oriented assessment, it was hardly surprising that the raters experienced problems

during the process of assessment. Despite the fact that the researcher had provided one training session and the guidelines of meaning-oriented assessment for the raters, disagreement among the results of assessment still occurred. This means that one training session and the guidelines might not be enough to provide the raters with sufficient scaffolding on the mechanisms and procedures of meaning-oriented assessment.

To sum up, two probable reasons for the disagreement among the results of the assessment by three raters could be concluded. Raters' lack of prior knowledge on SFL and the brief training session by the researcher might be the probable causes for the raters' problems when they implemented meaning-oriented assessment to assess the results of translation. As a result, the raters might have distinctive opinions on how to apply meaning-oriented assessment.

4.2 Inter-Rater Reliability

In Section 4.1, the similarities and differences of the results of assessment by all three raters have been discussed exhaustively. Based on those results, the researcher compiled and summarized the total of score deduction. The point deduction can be seen in the table below.

| Student | Total Deducted Points | | |
|----------------|------------------------------|----------------|----------------|
| | Rater 1 | Rater 2 | Rater 3 |
| Student 1 | 123.5 | 118 | 204.5 |
| Student 2 | 151.5 | 64.5 | 199 |
| Student 3 | 141.5 | 57 | 227 |
| Student 4 | 111 | 53 | 193.5 |
| Student 5 | 174.5 | 76.5 | 274.5 |
| Student 6 | 154 | 78 | 343 |

| | | | |
|------------|-------|------|-------|
| Student 7 | 164 | 68.5 | 298 |
| Student 8 | 128 | 69.5 | 328 |
| Student 9 | 131.5 | 59 | 293.5 |
| Student 10 | 104 | 58 | 239.5 |

Table 17. Summary of Point Deduction

In order to acquire each student's final score, 45 points (which is the maximum point in the meaning-oriented assessment) have to be deducted by the total deducted points in Table 17. The final scores can be seen in the table below.

| Student | Final Scores | | |
|------------|--------------|---------|---------|
| | Rater 1 | Rater 2 | Rater 3 |
| Student 1 | -78.5 | -73 | -159.5 |
| Student 2 | -106.5 | -19.5 | -154 |
| Student 3 | -96.5 | -12 | -182 |
| Student 4 | -66 | -8 | -148.5 |
| Student 5 | -129.5 | -31.5 | -229.5 |
| Student 6 | -109 | -33 | -298 |
| Student 7 | -119 | -23.5 | -253 |
| Student 8 | -83 | -24.5 | -283 |
| Student 9 | -86.5 | -14 | -248.5 |
| Student 10 | -59 | -13 | -194.5 |

Table 18. Summary of Final Scores

In order to test whether or not the assessment has inter-rater reliability, the researcher performed statistical tests by inputting the scores into SPSS. As mentioned in Chapter 3, firstly, the researcher conducted the normality test in order to find out the type of distribution (normal or abnormal). The result of the test can be seen in the table below.

| Tests of Normality | | | | | | | |
|--------------------|---------|---------------------------------|----|-------------------|--------------|----|------|
| Score | Rater | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
| | | Statistic | df | Sig. | Statistic | df | Sig. |
| | Rater 1 | .118 | 10 | .200 [*] | .976 | 10 | .943 |
| | Rater 2 | .239 | 10 | .112 | .778 | 10 | .008 |
| | Rater 3 | .146 | 10 | .200 [*] | .922 | 10 | .371 |

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Table 19. Result of Normality Test

There are two types of normality test as seen in the table above: Kolmogorov-Smirnov and Shapiro-Wilk. Thode (2002) mentioned that Kolmogorov-Smirnov test is used more often. However, Shapiro-Wilk is considered more appropriate for research with less than 50 data. Since this research utilized less than 50 data, the researcher, therefore, looked at the result generated from Shapiro-Wilk test in order to draw conclusion regarding the distribution.

In order to test the normality, the researcher formulated the null hypothesis (H_0) as “the data are normally distributed” and the alternative hypothesis (H_1) as “the data are not normally distributed.” To determine whether or not the null hypothesis is accepted or rejected, the researcher looked at the value of sig. (p-value) in Table 19. If the p-value is higher than 0.05, the null hypothesis is accepted, which means the data are normally distributed. On the other hand, if the p-value is lower than 0.05, the null hypothesis is rejected, and it can be concluded that the data are not normally distributed.

As seen in Table 19, the p-value of Rater 1 is 0.943, Rater 2 is 0.008, and Rater 3 is 0.371. In the case of Rater 1 and Rater 3, since the p-values are higher than 0.05, the null hypothesis is accepted, and this means that the data are normally

distributed. However, in the case of Rater 2, since the p-value is lower than 0.05, the null hypothesis is rejected, which means that the data are not normally distributed. Since not all the data are normally distributed, the researcher decided to employ non-parametric test, specifically Kruskal-Wallis test, as explained in Chapter 3.

Below is the result of Kruskal-Wallis test.

| | Score |
|-------------|--------|
| Chi-Square | 25.301 |
| df | 2 |
| Asymp. Sig. | .000 |

a. Kruskal Wallis Test
b. Grouping Variable: Rater

Table 20. Result of Kruskal-Wallis Test

As mentioned in Chapter 3, the null hypothesis (H_0) is “there is no significant difference across the scores by the raters,” and the alternative hypothesis (H_1) is “there is significant difference across the scores by the raters.” In order to find out whether the null hypothesis is accepted or rejected, the researcher looked at the value of Asymp. Sig. (p-value), which is 0.000. This means that the p-value is lower than 0.05; thus, the null hypothesis is rejected. It can then be concluded that there is significant difference across the scores by the raters and that there is low agreement among the three raters. Based on the results explained above, the researcher concluded that meaning-oriented assessment might not possess inter-rater reliability or possesses low level of inter-rater reliability.

4.3 Raters’ Perception on Meaning-oriented Assessment

This section discusses the perception of the raters after they were asked to implement meaning-oriented assessment to assess students' translation works. This section is divided into two sub-sections: sub-section 4.3.1 explains the strengths of meaning oriented assessment while sub-section 4.3.2 expounds the weaknesses of meaning-oriented assessment. As mentioned in Chapter 3, the analysis for this section comes from the interview conducted towards all three raters.

4.3.1 Strengths of Meaning-Oriented Assessment

4.3.1.1 Detailed criteria of assessment

All the raters agreed that meaning-oriented assessment presents detailed and comprehensive criteria in order to assess the translation works. This is apparent from how the raters described their experiences when conducting meaning-oriented assessment. Rater 3 mentioned:

“... [it is] very detailed because we have four criteria to undergo, [which are] experiential, logical, interpersonal, and textual, and each criteria also contains of word level, clause level, text level, and we should also undergo [analysis] on the accuracy and naturalness aspect. Not to mention the minor criteria. [It's] like they don't leave any hole available there.”

Similar to the opinions of Rater 3, Rater 1 also stated that meaning-oriented assessment is very detailed and comprehensive. Below is the response from Rater 1:

“... [it is] very detailed. I think it covers all there is to know about somebody's qualities in translation. Starting from the biggest unit, which is the text, until the smallest unit, which is word. And then it covers not only grammatical mistakes, but it also covers until whether the register is correct or not, whether the connection between clauses is correct or not. So I think it's very detailed. Very detailed.”

Rater 2 also expressed similar opinion regarding the degree of detail of meaning-oriented assessment. Rater 2 responded, "... it has been very detailed in exploring some components or elements of language. Although it is not holistically discussed, the components that the rater should assess have been detailed in explanation."

As seen in the quotes above, all raters mentioned the criteria of meaning-oriented assessment as the element that contribute to the level of detail of meaning-oriented assessment. Kim (2009) set up a range of criteria for meaning-oriented assessment. In order to reach decision of the error categories and point deductions, great effort has to be exerted and lengthy analysis has to be conducted.

Both Rater 1 and Rater 3 stated that they needed to categorize the type of errors (experiential, logical, interpersonal or textual metafunctional error); then, they needed to decide whether the error is a problem of accuracy or naturalness; afterwards, they must observe whether the error affects the translation in lexical, clausal, or text level; finally, they must judge the severity of the error by giving appropriate score deduction. The criteria, in their opinion, are highly detailed since the assessment observes every element available in Systemic Functional Linguistics.

Rater 2, on the other hand, believed that there are components that can be added or included in order to make meaning-oriented assessment more detailed and comprehensive. He mentioned:

"... there are some components [that can be included]. Components such as collocation, skewing, translation shifts, addition and omission, some elements by NAATI, such as too free or too literal translation, grammar, spelling, syntax, punctuation, or failure to finish translating."

Since meaning-oriented assessment emphasizes on meanings instead of surface-level linguistics, errors in grammatical level are placed under the category of Minor Errors, which warrants for 0.5 point deduction. Rater 2 believed that such errors should be expounded further and be made into their own independent criteria. Rater 2 mentioned several components of assessment for NAATI certification should be included into meaning-oriented assessment. This, however, contradicts Kim's (2009) purpose of setting up meaning-oriented assessment. Since the purpose of meaning-oriented assessment is to further complement and explain the criteria of NAATI certification (Kim, 2009), adding those undefined criteria of certification into meaning-oriented assessment would not be beneficial and would instead create confusion for the users of meaning-oriented assessment.

In brief, the raters view the detailed nature of meaning-oriented assessment as one of the advantages of the assessment. The degree of detail of meaning-oriented assessment ensures that even the smallest elements of translation to be included into the assessment.

4.3.1.2 High validity

Regarding the aspect of validity, all raters agree that meaning-oriented assessment is highly valid. This means that the assessment really examines what it is supposed to assess, which in this case is the quality of students' translation works. Rater 1 stated that, "[the assessment] is very valid. It covers everything. It assesses everything, starting from the textual level until the lexical level."

Rater 2 also mentioned similar notion as Rater 1. He explained that “meaning-oriented assessment focuses on the meaning, purpose and context of translation. Those elements are reflected in the criteria of meaning-oriented assessment itself.”

Furthermore, Rater 3 agreed that meaning-oriented assessment is highly valid, and it evaluates what it is supposed to evaluate. Below is the summary of Rater 3’s perception on the degree of validity of meaning oriented assessment.

“... if conducted in fit and proper condition, the result of the assessment would be very valid. It assesses a lot of criteria and sub-items, and there are a lot of steps conducted in order to find out whether certain error is a problem of accuracy or naturalness, whether it is a problem of lexis, clause, or text. I think meaning-oriented assessment is very valid because it is complete and comprehensive, and it ensures that there is no hole for the accuracy and naturalness problems.”

Based on the quotes above, the researcher concluded that all the raters believed the degree of validity is somehow interconnected to the degree of detail of meaning-oriented assessment. When the raters were asked for further elaboration on the issue of validity, they related their reasoning to the matter of details. This means the raters perceived that the amount of details in meaning-oriented assessment contributes to the level of validity of the assessment. The more detailed the assessment is, the more valid it is.

Interestingly, Rater 2 stated that meaning-oriented assessment pays attention to the “meaning, purpose and context of translation” when he was asked to elaborate more about the aspect of validity of meaning-oriented assessment. Indeed, it is important that the translation communicates the meanings and fulfills the purpose of the original, and it should be conducted by considering the contextual situations surrounding the act of translation. This, in his opinion, further enhances the validity

of meaning-oriented assessment since the act of translation should always be conducted based on those aspects.

To sum up, all the raters believed that validity is another advantage of meaning-oriented assessment. According to the recurring statements by the raters, meaning-oriented assessment examines what it is supposed to examine, which in this case is the quality of translation conducted by the students.

4.3.2 Weaknesses of Meaning-Oriented Assessment

4.3.2.1 Time inefficiency

All the raters agree that meaning-oriented assessment is time-consuming. The raters admitted that the time spent on assessing the students' translation works is quite lengthy, and it takes minimum 30 minutes to maximum one and a half hour to evaluate one student's translation work. This, in the raters' general opinion, affects other aspects of meaning-oriented assessment, particularly the efficiency and the practicality of the assessment.

Rater 1 believed since meaning-oriented assessment is time-consuming, it would be more appropriate to implement meaning-oriented assessment for the purpose of professional certification instead of for assessing students' translation works. Rater 1 perceived that it does not worth the time if the raters or translator trainers spend 30 minutes to one and half hour only to deal with one student's translation works. This is evident from the statement by Rater 1 below.

“It is time-consuming. If it is for a student, I think one and a half hour only to check on one paper is too long, but if it is for certification or professional purpose, to assess somebody's ability whether or not the person can produce quality

translation or not, I think it is worth the time, but if it is only for students, I think one and a half hour is too long. So if you are asking whether this [assessment] is time-efficient or not, if for students, [it is] not. But for [the assessment for] professional translators, I think it is time-efficient. One and a half hour is good.”

Rater 2 also mentioned similar issue. Compared to the method of assessment that he often uses in the classroom, meaning-oriented assessment is more time-consuming, especially if it is the rater’s first attempt to utilize meaning-oriented assessment. He claimed that when the rater is accustomed to conducting assessment using meaning-oriented assessment, the process might be less tedious and the time spent in assessing would be much shorter. However, considering the amount of the students in translation class and the limited time given by the university to the lecturers or translator trainers in assessing students’ translation works, Rater 2 thought it would be impractical to implement meaning-oriented assessment in the translation classrooms.

Rater 3 also shared the same opinion with Rater 1 and Rater 2. However, Rater 3 suspected the probable reason for the time inefficiency was due to the amount of translation items that need to be assessed. Rater 3 also mentioned that he sometimes lost track of his assessment, so he needed to refer back to the previous evaluation constantly in order to ensure the consistency of the assessment. As a result, the time spent on assessing the translation works became longer. Rater 3 also mentioned that the implementation of meaning-oriented assessment in the translation classroom might tax or burden the lecturers. By implementing meaning-oriented assessment, the lecturers would need to dedicate a large portion of their time only to

assess the students' translations, which according to Rater 3 is illogical since lecturers have other responsibilities besides assessing the students' translation works.

In brief, the time consumption is considered as one of the weaknesses of meaning-oriented assessment. The raters view the amount of time spent on evaluating students' translation works using meaning-oriented assessment is disproportional to the amount of other works that lecturers have to do.

4.3.2.2 Complicated mechanism

Based on the dominant views, it can be concluded that the mechanism of meaning-oriented assessment is difficult to utilize. The raters, once again, attributed this weakness to the matter of detail of meaning-oriented assessment. The raters generally believed that one of the strengths of meaning-oriented assessment is the detailed criteria of assessment. However, because it is very detailed, the raters found difficulties in utilizing meaning-oriented assessment. As mentioned before, there are several stages of analysis that the raters need to do before they decide the amount of points that they need to deduct. The process of undergoing these stages, in the raters' opinions, is long and tedious. This can be seen from the summary of statement by Rater 1 below.

“... the criteria for the assessment is definitely more complicated than the [assessment] that I have done. It's very detailed. There are a lot of criteria that need to be looked at. I think for translation students, it's too complicated. It's not easy for me to assess.”

Rater 1 particularly experienced difficulty when giving the point deduction. As seen in the criteria of meaning-oriented assessment, there are ranges of scores for each

type of error. There are range of 1 to 2, 1 to 3, and 3 to 5 points deduction. Rater 1 found the process of assigning point deduction easier if the range of point deduction only includes two options. For example, experiential error in terms of accuracy in the level of lexis warrants for 1 to 2 points deduction. Rater 1 could immediately assign less severe error with 1 point deduction and more severe error with 2 point deduction. However, Rater 1 stated that she found difficulties deciding when the range of point deduction includes three options, such 3 to 5 points deduction. Rater 1 mentioned that she had no problem assigning 3 (for less severe error) and 5 (for more severe error) points deduction, but she could not decide when to assign 4 points deduction. The following is the summary of her statement.

“... if it's only two [options], you can just see that if it's low in the band, it means the mistake is slight, while here the mistake is severe. You can see the difference. But if for example, if the range is a bit high, like three-four-five, then you need to decide when you put the middle [point], and that is more difficult. But if it's only one to two, it's easy. Slight difference [in meaning should be given] one [point deduction] and big difference [should be given] two. But three-four-five? It is difficult to see. I normally just use either three or five instead of using four.”

Rater 3 also shared the same opinion with Rater 1. Rater 3 compared meaning-oriented assessment with the method of assessment he usually uses in the classroom, the so-called CAN (Clarity, Accuracy, and Naturalness) method, which is derived from Larson (1984). Unlike Rater 1 who experienced difficulties in assigning points deduction, Rater 3 did not find it difficult to assign point deduction for each error. However, he expressed his confusion on the existing range of point deduction. Kim (2009) had determined the scoring range for each criterion in meaning-oriented assessment, and each criterion has different range of point deduction. Rater 3 admitted that the different range or scale of point deduction sometimes confused him

when he was in the process of assessing. Below is the summary of Rater 3's response regarding this matter.

"I think [the scoring range] is still quite okay although my question is why in several aspects, [the range] is one and two and then in another criteria, it starts from three to five. So once I made a mistake. I put two [for the point deduction] and then I realized that three is the minimum [point deduction] in this case, so I have to adjust some scores that I have given. So I think it's better if we use the same scale, one to two or one to three. [...] I don't know about the reason [why the scoring range is designed like that]. There must be some explanation about that."

Rater 3 also found difficulties when he had to classify and label the errors based on the criteria of meaning-oriented assessment. For example, when he found erroneous translation, he needed to refer back to meaning-oriented guide and his previous assessment in order to decide whether the error belongs to experiential, logical, interpersonal or textual metafunction. This, in the end, caused Rater 3 to spend longer time in conducting the assessment. Rater 3 stated, "... [assigning point deduction] is not a big thing. It's not a big challenge. The biggest challenge is to identify whether this is logical, whether this is experiential, and so on."

In brief, the complicated mechanism is viewed as another weakness of meaning-oriented assessment. According to the predominant views from the raters, the error classification and range of point deduction contribute to the complication in the usage of meaning-oriented assessment. Some of the raters mentioned that the detailed nature of the meaning-oriented assessment leads to the complicated mechanism.

4.3.2.3 Subjective interference

The issue of subjectivity was repeatedly mentioned during the course of interview. All of the three raters agree that there are elements in meaning-oriented assessment that can cause subjectivity. Ideally, a method of assessment should produce objective results and minimize the subjectivity of the assessors. However, the raters particularly underlined that subjectivity could interfere during the process of assessment using meaning-oriented assessment.

Rater 1 believed that when the process of checking and editing the results of translation is conducted, subjectivity always comes into play. One rater or assessor might perceive one particular word as erroneous translation whereas another rater or assessor might not think the same translation as an error. She also emphasized that “editing is never going to be very objective,” yet since the criteria of meaning-oriented assessment has been set up and explained, the subjectivity can be minimized. This is apparent from the quote by Rater 1 below.

“... editing is never going to be very objective, but this subjectivity is reduced by the criteria that you provide. This will create more objective result, but I think it’s not going to be a hundred percent objective. I think it’s going to be somewhere in between forty five to maybe sixty percent of objectivity.”

Rater 2 also shared similar opinion as Rater 1. According to Rater 2, subjectivity may occur during the assessment when the raters found cases of errors that are not stated or included in the criteria of meaning-oriented assessment. Rater 2 gave cases of mistranslation and omission of collocation as examples. According to Rater 2, since those errors are not included in the criteria, raters need to employ their subjective opinions in order to give appropriate point deduction. Furthermore, Rater 2 mentioned that subjectivity can happen during the assignment of point deduction.

Nevertheless, similar to Rater 1, Rater 2 mentioned that the criteria could help balance the subjectivity and boost the objectivity of the assessment. As long as the raters follow the criteria of meaning-oriented assessment closely, the results of assessment could be fairer and more objective.

Rater 3 has slightly distinctive view regarding the matter of subjectivity and objectivity. He stated that the objectivity and subjectivity are dependent on the raters' situations. The criteria have been set up and organized well and comprehensively. However, when the raters are not in good condition, it would be difficult to achieve objective results of assessment. Below is the response from Rater 3.

“I think if conducted in comfortable or suitable or ideal condition, the method could generate objective result. [...] If it is the other way around, if it is conducted in a not-so-ideal [situation], the subjectivity will occur. Definitely. The system is very well, perfect, comprehensive, but it will depend on the raters' condition.”

To sum up, despite the fact that all raters have different arguments regarding the forms and causes of subjectivity in meaning-oriented assessment, all of them stated that there is high possibility for subjectivity to occur during the process of assessment. As a result, all raters perceived subjectivity as one of the weaknesses of meaning-oriented assessment.

4.3.2.4 Low reliability

Based on dominant views, the issue of reliability is perceived as a weakness of meaning-oriented assessment. 2 out of 3 raters believed that in the case of meaning-oriented assessment, there is a possibility that the raters would produce very distinctive results from one another; thus, reliability would be difficult to achieve.

The excerpt below is the response from Rater 1 when she was asked about the reliability of meaning-oriented assessment.

“In my experience, when you edit somebody’s work, other editor or the second editor will produce a different result. [...] If you’re asking whether this [assessment] is reliable or not in terms of whether they will produce the same score, I don’t think so. Something that I consider correct might be considered wrong by other translators. [...] So I think based on this fact, I think it’s going to be difficult for you to score a very good reliability. If you’re going to assess the reliability, it’s going to be very low.”

As seen above, Rater 1 related the low reliability to the issue of subjectivity of the raters, stating that each rater has different perceptions and opinions about errors. Since each rater is different, he or she would produce different results of assessment as well. The difference of the results would affect the reliability level of meaning-oriented assessment.

Rater 3 also expressed his uncertainty regarding the reliability of meaning-oriented assessment. Besides the fact that each rater has different perception, Rater 3 also mentioned that the raters’ lack of intra-rater reliability might cause a problem. Intra-rater reliability here includes the raters’ situation, which includes focus and endurance, while conducting the assessment. Rater 3 stated that raters for meaning-oriented assessment have to be physically and mentally fit. This is because the process of meaning-oriented assessment is long, and the items of assessment are “rich in terms of variety.” Below is the summary of Rater 3’s responses in terms of reliability.

“The first thing is the different perception [among raters]. The second thing is the endurance of the raters. I think in doing this, the raters should have quite good physical condition because the items [of assessment] are so rich in terms of variety. If you lose the focus, you will lose the mood. [...] Psychologically, it could be a kind of burden, so you need to maintain your focus well. [...] I think if the

raters keep pushing themselves to do [the assessment] while actually they are already tired, it may affect the reliability for sure.”

As seen in the excerpt above, Rater 3 broadened the understanding of reliability. Not only did he consider the reliability among raters, he also thought about the reliability within the self of the raters. He believed that the raters would not be able to properly maintain the consistency of their own assessment if they are not in fit situation.

Rater 2, however, has slightly different opinion. Rater 2 believed the results of assessment would be reliable due to the existing criteria in meaning-oriented assessment, but the final scores might be different. He believed that the criteria of meaning-oriented assessment would help the raters in identifying and labeling the type of metafunctional errors. If the raters follow the description of each metafunctional error closely, Rater 2 believed that in the end, the raters will arrive at the same conclusion when they identify the errors and categorize them based on the criteria of meaning-oriented assessment. However, in the stage of assigning point deduction, this could be different from one rater to another. Two raters might be able to spot the same error and categorize it under the same criteria of metafunctional errors, but it is unlikely that they will give the same point deduction since there are ranges of point deduction in meaning-oriented assessment. Hence, the final score given by different raters might not be the same despite the fact that both of them assess the same translation works.

To sum up, when the raters were asked about the level of reliability of meaning-oriented assessment, all three of them could not confidently answer that the results of the assessment would be reliable if compared among three raters. Thus, the

low reliability is considered as one of the weaknesses of meaning-oriented assessment. However, the researcher also needs to caution the readers that the low reliability might not be caused by the system of meaning-oriented assessment itself. As mentioned previously, even though the criteria and the system of assessment had been outlined, the raters' different experiences in dealing with errors in translation might indirectly cause the differences in the results of assessment, which leads to the low level of reliability. Rater 3 also mentioned random factors might also affect the reliability of assessment such as fatigue, unfit condition, psychological burden, and so on, and this is also in line with the claim by Bachman (1990).

4.3.2.5 Problematic implementation

During the interview process, the researcher asked about the possibility for the implementation of meaning-oriented assessment in formative and summative assessment. Based on dominant views, it could be concluded that the implementation of meaning-oriented assessment in formative and summative assessment would be problematic. All raters have different views regarding this matter, but most of them suggested modification and adjustment for meaning-oriented assessment should it be used for formative and summative purposes.

Rater 1 reasoned that due to the detailed nature of meaning-oriented assessment, it is impractical and time-consuming to use meaning-oriented assessment for formative purposes, such as assessing translation assignments or quizzes. She stated that "it's going to be difficult for the lecturer or the assessor to do this, unless some criteria are taken out." In other words, meaning-oriented assessment could be

used if some criteria were omitted as an effort to simplify the assessment. This, however, would mean that the result of the assessment would not be as detailed as the result produced with original assessment. Similarly, Rater 1 believed that with slight modification, meaning-oriented assessment can be used for summative purposes. For example, instead of looking at every element in the translation, the lecturer can check the overall or the general impression of the translation. Moreover, she suggested omitting the criteria of naturalness since the students, in her opinion, need to focus on the aspect of accuracy first before tackling the problem of naturalness.

Rater 2 believed that meaning-oriented assessment would be beneficial if it was implemented for both formative and summative purposes. He claimed that meaning-oriented assessment would be helpful and useful for both the lecturers and the students in their teaching and learning process since it is very detailed. Nevertheless, since the process is very time-consuming, Rater 2 mentioned that the lecturers must adjust their schedules and manage their time well. Considering the amount of lecturers' tasks, he believed with proper time and workload management, meaning-oriented assessment could be implemented successfully in both formative and summative assessment.

Rater 3, however, was not as optimistic as Rater 2. The conclusion of his arguments shows that meaning-oriented assessment is not suitable to be used for formative purposes. He mentioned that the process of assessment is too time-consuming and repetitive, and it requires much focus and energy. If the lecturers were dealing with a very large translation class, they would need to prepare every single step so that the assessment could be implemented successfully. This, according to

Rater 3, would burden the lecturers. On the other hand, Rater 3 believed that it is acceptable to implement meaning-oriented assessment for summative purposes. Since summative assessment evaluates “the whole process of learning in one semester,” Rater 3 thought that meaning-oriented assessment could be utilized in order to evaluate whether the students have successfully achieved the goals of learning and are able to translate naturally and accurately.

4.4 Result of Triangulation

Different topics related to meaning-oriented assessment have been discussed extensively in the previous sections. The results of analysis explained in the previous sections were generated using different methods, and it is important to integrate those results in order to draw a final conclusion. The usage of mixed methods in this case enabled the researcher to triangulate the research findings in order to provide final remarks on meaning-oriented assessment in assessing students’ translation works. Below is the illustration of the triangulation process.

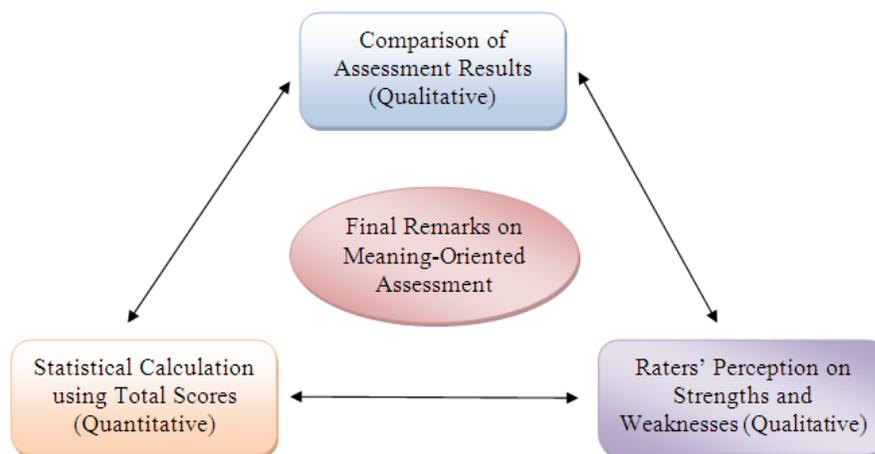


Figure 7. Illustration of Triangulation Process

From the comparison across assessments conducted by Rater 1, Rater 2, and Rater 3, it could be concluded that there were similarities and differences across the findings. Several patterns of similar cases were recurring across the data, and the patterns extracted from the data revealed that the differences among the raters' assessment were more evident than the similarities. The differences can be seen in terms of error categories and assignment of point deduction. A descriptive view on the results of assessment by the raters shows high degree of inconsistency and disagreement across the raters. However, the comparison across the results of assessment is qualitative in nature, and it generates subjective interpretation from the researcher. Moreover, the result of comparison does not reveal information on the inter-rater reliability of meaning-oriented assessment. Therefore, quantitative measure was needed in order to shed light on the question of the inter-rater reliability.

In this research, the quantitative aspect would support the qualitative aspect of the research by providing more concrete information on the notion of inter-rater reliability. As mentioned before, the total scores by the raters were inputted into SPSS, and the hypotheses were tested. The result of the SPSS calculation shows that there was significant difference among the results of assessment. This was in line with the result of comparison mentioned before in which the results of the assessment by three different raters tended to be distinctive in nature. Since there is significant difference among the results of assessment, it means that meaning-oriented assessment either has low inter-rater reliability or no inter-rater reliability.

After employing the qualitative and quantitative measures, the researcher also added information gathered through another qualitative means, which was the

interview, in order to strengthen the arguments further. As explained above, the purpose of the interview was to elicit information regarding the raters' perception on meaning-oriented assessment, specifically the strengths and weaknesses of the assessment. The result revealed that the weaknesses were more dominant than the strengths. Among the weaknesses mentioned by the raters, the issue of low reliability was also discussed, and this further confirmed the results of the comparison of assessment and the statistical calculation.

As mentioned before, all of these findings were triangulated in order to provide a final remark regarding the usage of meaning-oriented assessment in assessing students' translation works. In other words, among the data generated from this study, it was revealed that the meaning-oriented assessment was not reliable. This can be seen from the great differences in the results of assessment, the result of statistical calculation, and the opinions of the raters. The assessment conducted by three different raters revealed inconsistent result, both from the descriptive perspective and calculation, and based on the analysis on interviews, the raters believed that the meaning-oriented assessment would produce results with low reliability due to raters' different views and other random factors. Based on the result of three analyses, therefore, the researcher concluded that it might be difficult to implement meaning-oriented assessment in assessing students' translation works. As a result, the researcher suggested modification or proper and sufficient training process for any parties who would like to implement meaning-oriented assessment.

CHAPTER 5

CONCLUSION AND SUGGESTIONS

This chapter summarizes and draws some conclusions from the findings of the research. The answers to the research questions and further implication of the research would be discussed. Moreover, it also provides suggestions for the future researchers who wish to conduct similar research.

5.1. Conclusion

This research specifically aims at discovering the implementation of meaning-oriented assessment in assessing students' translation works, particularly on the results of assessment by the raters, the inter-rater reliability of meaning-oriented assessment, and strengths and weaknesses of meaning-oriented assessment.

The overall comparison of the results of assessment which was conducted by three raters have shown that the differences are more dominant than the similarities. The differences occurred in terms of error categorization and assignment of point deduction. There were cases in which the raters categorized the errors into the same category and assigned the same point deduction. However, in a glance, it can be seen that the number of such occurrence is significantly lower than the number of cases with different categorization and point deduction. To answer the first research question, the researcher, therefore, concluded that the results of the assessment conducted by three different raters generated different results despite the fact that there were some similarities in the comparison.

To support the findings from the comparison of assessment results and to answer the second research question, the researcher employed statistical calculation in order to find out the inter-rater reliability of the assessment. The calculation utilizing SPSS and Kruskal-Wallis test revealed that there is significance difference among the results of assessment. This means that the raters had different perspectives on how to implement meaning-oriented assessment and that the results generated by three raters using meaning-oriented assessment were not consistent. In other words, it could be concluded that meaning-oriented assessment either possesses low level of inter-rater reliability or does not possess inter-rater reliability. This further solidifies the findings acquired from the comparison of assessment results which was previously discussed.

In order to find out the raters' perception about meaning-oriented assessment and to answer the third research question, the researcher interviewed the three raters. From the interview, the raters believed that meaning-oriented assessment provided detailed criteria for the assessment. It was very comprehensive, and it accounted all types of errors, both major and minor. Furthermore, all raters also believed that meaning-oriented assessment evaluated what it was supposed to evaluate. In other words, the assessment is of high validity. On the other hand, the raters also thought that due to the detailed criteria, the assessment was too time-consuming and too complicated. Moreover, there was a possibility for subjectivity to play in the assessment, and the raters also predicted that the degree of reliability would be very low due to the raters' different experiences and views regarding the errors and due to random factors such as fatigue, unfit condition, and so on. Finally, the raters also

believed that the implementation of meaning-oriented assessment for both formative and summative purposes would be problematic.

Based on the results found in this research, the researcher triangulated those results in order to draw final remarks regarding meaning-oriented assessment. The final conclusion that the researcher could draw was that the implementation of meaning-oriented assessment in assessing students' translation works might be difficult. This result, however, could not be generalized to all cases of meaning-oriented assessment. As mentioned in the limitation of study, this research could be categorized as a small-scale study which involved small number of participants. As a consequence, the results generated by this study should not be considered as representative to similar studies. Moreover, it is important to remember that there are many factors that may influence the reliability of this assessment, including the raters' differences in experiences and views and other random factors such as stress, preferences, fatigue, and so on. In addition, as far as the researcher's knowledge, there were no previous studies which analyzed the inter-rater reliability, strengths, and weaknesses of meaning-oriented assessment. As a result, the researcher was unable to compare the results of this research with the results from previous studies.

For the translator trainers who are considering to implement meaning-oriented assessment, it is suggested to consider the possibility of modifying the assessment criteria or maximizing the training process of implementing meaning-oriented assessment. It is hoped that by doing the actions mentioned previously, the weaknesses of meaning-oriented assessment can be mitigated, and the strengths can be retained.

5.2. Suggestions

For future researchers who are interested in exploring the implementation of meaning-oriented assessment, research with larger participants and more raters might give more interesting and significant findings. As explained previously, this research only gathered data from ten students and employed participation from three raters. If the research could be conducted in larger scale, it would probably generate distinctive and diverse results as opposed to the results of this research.

Furthermore, the researcher also suggested the future researchers to maximize the training or scaffolding process for the raters in order to ensure that the raters comprehend the whole implementation of meaning-oriented assessment. It would also assist the future researchers if they could find raters with previous knowledge on Systemic Functional Linguistics. As mentioned in the previous chapters, all raters who participated in this research had no prior knowledge about Systemic Functional Linguistics and meaning-oriented assessment. Moreover, the training or the scaffolding process that the raters went through was also too short. By compensating these shortcomings, the future researchers might find more consistent results in their research.

REFERENCES

- Alwi, H. (1992). *Modalitas dalam bahasa Indonesia* (Modality in bahasa Indonesia) (Doctoral dissertation). Yogyakarta: Kanisius.
- Angelelli, C. V. (2009). Using a rubric to assess translation ability. In C. V. Angelelli, & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: a call for dialogue between research and practice* (pp. 13-47). Philadelphia, PA: John Benjamins.
- Angelelli, C. V., & Jacobson, H. E. (2009). Introduction: Testing and assessment in translation and interpreting studies: a call for dialogue between research and practice. In C. V. Angelelli, & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: a call for dialogue between research and practice* (pp. 1-10). Philadelphia, PA: John Benjamins.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, OX: Oxford University Press.
- Baer, B. J., & Bystrova-McIntyre, T. (2009). Assessing cohesion: Developing assessment tools on the basis of comparable corpora. In C. V. Angelelli, & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: a call for dialogue between research and practice* (pp. 159-184). Philadelphia, PA: John Benjamins.
- Bowker, L. (2000). A corpus-based approach to evaluating student translations. *Evaluation and Translation: Special Issue of the Translator*, 6(2), 183-210.
- Bowker, L. (2001). Towards a methodology for a corpus-based approach to translation evaluation. *Meta: journal des traducteurs / Meta: Translators' Journal*, 46, 345-364.
- Bowker, L. (2003). Corpus based application for translator training: Exploring the possibilities. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based approaches to contrastive linguistics and translation studies* (pp. 169-183). New York, NY: Rodopi.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York, NY: Pearson Education.
- Caffarel, A., Martin, J. R., & Matthiessen, C. M. I. M. (Eds.). (2004). *Language typology: A functional perspective*. Philadelphia, PA: John Benjamins.
- Colina, S. (2003). *Translation teaching: From research to the classroom: A handbook for teachers*. Boston, MA: McGraw-Hill.

- Colina, S. (2011). Evaluation/Assessment. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies: Volume 2* (pp. 43-48). Philadelphia: John Benjamins.
- Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians*. New Jersey, NJ: John Wiley & Sons.
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed method approaches* (4th edition). Los Angeles, LA: SAGE Publications.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford, OX: Oxford University Press.
- Eggins, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). New York, NY: Continuum.
- Flesch, R. (n.d.). *How to write plain English*. Retrieved from http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring extent of agreement among raters* (4th ed.). Gaithersburg, MD: Advanced Analytics, LLC.
- Halliday, M. A. K. (1992). Language theory and translation practice. *Rivista Internazionale di Tecnica Della Traduzione*, 15-25.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Oxon, OX: Hodder Arnold.
- Harris, D. P. (1969). *Testing English as a second language*. New Delhi, ND: Tata McGraw-Hill.
- Hatim, B., & Mason, I. (1997). *The translator as communicator*. New York, NY: Routledge.
- Heyvaert, M., Maes, B., & Onghena, P. (2011). Applying mixed methods research at the synthesis level: An overview. *Research in the Schools*, 18(1), 12-24.
- Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, 17, 1-10.
- Jayanti, R. R. (2012). *Modalization and modulation expressions of the English-Bahasa Indonesia "A Beautiful Mind" movie texts* (Undergraduate thesis). Yogyakarta: Universitas Negeri Yogyakarta.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.

- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130-144.
- Karoly, K. (2006). *Translation studies: The analysis of translation as a text* [Powerpoint Slides]. Retrieved from seas3.elte.hu/coursematerial/**Karoly**Krisztina/TS-lecture6.ppt.
- Khanmohammad, H., & Osanloo, M. (2009). Moving toward objective scoring: A rubric for translation assessment. *JELS*, 1(1), 131-153.
- Kim, M. (2007) Using systemic functional text analysis for translator education. *The Interpreter and Translator Trainer*, 1(2), 223-246. DOI: 10.1080/1750399X.2007.10798759.
- Kim, M. (2009). Meaning-oriented assessment of translations: SFL and its application to formative assessment. In C. V. Angelelli, & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: a call for dialogue between research and practice* (pp. 123-158). Philadelphia, PA: John Benjamins.
- Kim, M. (2010). Translation error analysis: A systemic functional grammar approach, In C. Coffin, T. Lillis, & K. O'Halloran (Eds.), *Applied linguistics methods: A reader* (pp. 84-94). New York, NY: Routledge.
- Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel* (No. RBR-8-75). Naval Technical Training Command Millington TN Research Branch.
- Kubiszyn, T., & Borich, G. D. (2013). *Educational testing & measurement: Classroom application and practice* (10th ed.). New Jersey, NJ: John Wiley & Sons.
- Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet? *Evaluation and Translation: Special Issue of the Translator*, 6(2), 149-168.
- Larson, M. (1984). *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.
- Leech, B. L. (2002). Asking questions: Techniques for semistructured interview. *PS: Political Science and Politics*, 35(4), 665-668.
- Manfredi, M. (2011). Systemic functional linguistics as a tool for translation teaching: Towards a meaningful practice. *Rivista Internazionale di Tecnica Della Traduzione*, 13, 49-62.

- Martin, J. (2010). Language, register and genre. In C. Coffin, T. Lillis, & K. O'Halloran (Eds.), *Applied linguistics methods: A reader* (pp. 12-32). New York, NY: Routledge.
- Martin, J. R., Matthiessen, C. M. I. M., & Painter, C. (2010). *Deploying functional grammar*. Beijing: The Commercial Press.
- Melis, N. M., & Albir, A. H. (2001). Assessment in translation studies: Research needs. *Meta: journal des traducteurs / Meta: Translators' Journal*, 46, 272-287.
- Moskal, B. M. (2000). Scoring rubrics: what, when, and how? *Practical Assessment, Research, and Evaluation*, 7(3). Retrieved September 15, 2015 from <http://pareonline.net/getvn.asp?v=7&n=3>.
- Munday, J. (2008). *Introducing translation studies: theories and applications* (2nd ed.). New York, NY: Routledge.
- Newmark, P. (1987). The use of systemic linguistics in translation analysis and criticism. In R. Steele, & T. Threadgold (Eds.), *Language topics: Essays in honor of Michael Halliday* (pp. 293-304). Philadelphia, PA: John Benjamins.
- Pearson, J. (2003). Using parallel texts in the translator training environment. In F. Zanettin, S. Bernardini, & D. Stewart (Eds.), *Corpora in translator education* (pp. 15-24). Manchester: St. Jerome Publishing.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of English language*. New York, NY: Longman.
- Sainz, M. J. (1994). Student-centred corrections of translations. In C. Dollerup, & A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions* (pp. 133-142). Philadelphia, PA: John Benjamins.
- Saldanha, G., & O'Brien, S. (2014). *Research methodologies in translation studies*. New York, NY: Routledge.
- Shiyab, S. M. (2013). Translation quality assessment: A perspective on pedagogy. *Arab World English Journal: Special Issues on Translation*, 2, 42-50.
- Sutrisno, A. N. (2014). *Knowledge transfer through dual degree programs: Perspectives from Indonesian universities* (doctoral dissertation). Queensland University of Technology, Queensland, Australia.
- Taufik, A. (2014). Analysing the reliability of functional-pragmatic model marking scheme in translation assessment. *Kongress Internasional Masyarakat Linguistik Indonesia: Peran Bahasa Ibu dan Bahasa Nasional dalam Pengembangan Potensi Penutur Bahasa* (pp. 24-29). Bandar Lampung: Masyarakat Linguistik Indonesia.

- The Editorial Board (2016, February 26). Don't Let Drugs Rule the Nation. *The Jakarta Post*. Retrieved from <http://www.thejakartapost.com/news/2016/02/26/editorial-don-t-let-drugs-rule-nation.html>.
- The Editorial Board (2016, March 15). Outrage over an Italian Student's Murder in Egypt. *The New York Times*. Retrieved from <http://www.nytimes.com/2016/02/11/opinion/outrage-over-an-italian-students-murder-in-egypt.html>.
- Thode, H. C. (2002). *Testing for normality*. New York, NY: Marcel-Dekker, Inc.
- Thompson, G. (2014). *Introducing functional grammar* (3rd ed.). New York, NY: Routledge.
- Waddington, C. (2001a). Should translations be assessed holistically or through error analysis? *Hermes, Journal of Linguistics*, 26, 15-37.
- Waddington, C. (2001b). Different methods of evaluating student translations: The question of validity. *Meta: journal des traducteurs / Meta: Translators' Journal*, 46(2), 311-325.
- Williams, M. (2009). Translation quality assessment. *Mutatis Mutandis*, 2(1), 3-23.

APPENDIX 1 SOURCE TEXT

Don't Let Drugs Rule the Nation

President Joko “Jokowi” Widodo’s commitment to the fight against drugs is part of his responsibility to protect the nation. Last year he repeatedly warned of “a drug emergency”, saying 50 people died in relation to drug use every day and 4.5 million people were addicted, giving justification to execute drug convicts.

At the opening of a Cabinet meeting on Wednesday, the President said he wanted “more intensive, braver, crazier and more comprehensive integrated efforts to eradicate drugs”. During the meeting, which came on the heels of a series of raids that led to the arrest of military and police personnel and, reportedly, a House of Representatives lawmaker, Jokowi declared a war on drugs.

Whether the braver and crazier measures will include another round of executions of drug convicts remains to be seen. But the way the government and law enforcers have combatted drugs so far has not shifted from the textbook: Raids are conducted here and there, most of the time with much fanfare, dozens are arrested and it is business as usual again until the next cycle of raids.

Worse, law enforcement against drug crimes is limited to small fry and is unable to catch the big fish, who use their money or influence to evade justice. Indeed the drug business is so secretive that law enforcers may find it difficult to capture the kingpins, but after so many raids and tons of drugs have been discovered, the public has yet to hear of any drug ring leader being arrested and prosecuted.

Challenges facing drug eradication also include impunity or at least light punishments. Many would remember an Army officer, a general’s son, who was found guilty of drug abuse more than 15 years ago but had his prison sentence reduced and his dismissal from the military revoked. He has won promotions since then.

Abuse of power is another cause for concern, especially in the antidrug campaign. This ranges from reports of people without any history of drug use being caught up in drug checks and forced to pay huge sums of money for their release, as well as law enforcers who act as drug dealers or protect drug rings.

Now that the President has called for integrated action to win the war on drugs, the National Narcotics Agency needs to draw up a clear strategy and set concrete targets. The agency must have identified hotspots where drugs are trafficked and circulated, as well as the people who are involved in the drug supply chain.

Tightening supervision of prisons, which the President believes are centers of drug trafficking, is mandatory. Random checks are a must — every day if possible, rather than every month as suggested by the President — given the scale of drug abuse facing the country.

The magnitude of the national campaign against drugs may require a powerful agency like the US Drug Enforcement Administration. A debate over this possibility should start sooner rather than later, simply because the war on drugs has already begun.

Taken from <http://www.thejakartapost.com/news/2016/02/26/editorial-don-t-let-drugs-rule-nation.html>

APPENDIX 2 INTERVIEW PROTOCOL

Title of Research : Meaning-oriented Assessment: An Analysis on Inter-rater Reliability, Strengths, and Weaknesses
Date & Time :
Place of Interview :
Interviewee :

Good morning / afternoon.

This interview is conducted as a part of my master study in the Faculty of English Applied Linguistics at Atma Jaya Catholic University. The topic of this research is related to the notion of translation assessment in translation classroom. There are three aims of this research. First is to find out the result of assessment which is conducted using meaning-oriented assessment by the raters. Second is to discover the reliability of meaning-oriented assessment. Finally, this research tries to identify the raters' perceptions about meaning-oriented assessment. This interview will provide insights and information to achieve the third aim.

Your identity as the rater for this project and as the source for this interview will be treated as confidential, and you are free to withdraw from this project should you wish to. This interview will be recorded using a recording device. The audio recording will later on be transcribed, and the transcription will be sent to you afterwards so that you will be able to verify your responses before it is used for the research process. The recording itself will be stored in the place where only the researcher can access, and it will be properly disposed five years after the researcher finalizes her research process. This interview will take place for approximately an hour.

We are going to begin shortly (*test the recording device*).

-
1. When teaching translation subjects, which method of assessment do you usually use in order to evaluate the students' translation works?
 - Can you please explain more about the method of assessment that you mentioned?
 - What do you think about the method of assessment?
 - What do your students think about the method of assessment?
 2. Could you describe your experience in assessing the students' translation works using meaning-oriented assessment?
 - How much time did you need to complete one assessment?

- What do you think about the procedures of error identification?
 - What do you think about the error assignment?
 - Can you please describe your experiences in terms of score assignment?
3. What do you think about the guide to meaning-oriented assessment?
 - Is it complete / comprehensive enough?
 - Does it answer or solve your confusion? Why do you think so?
 - Does it give clear explanation about the procedures of meaning-oriented assessment?
 4. In your opinion, what are the strengths and weaknesses of the assessment? Let us start with the strengths.
 - Could you please elaborate more on the aspect of ...?
 - Can you give examples for ... point?
 - What are other advantages that you can think of?
 5. In contrast, what are the weaknesses of meaning-oriented assessment?
 - Could you please elaborate more on the aspect of ...?
 - Can you give examples for ... point?
 - What are other disadvantages that you can think of?
 6. In your view, how was the level of reliability of meaning-oriented assessment? Reliability here means whether or not the other raters would produce similar results as you did.
 - Why do you think it is reliable / unreliable?
 - Can you give examples for ... point?
 7. In your view, how was the level of validity of meaning-oriented assessment? Validity here means whether or not this assessment really examines what it is supposed to be assessed.
 - Why do you think it is valid / invalid?
 - Can you give examples for ... point?
 8. In your view, how was the level of systematicity of meaning-oriented assessment? Systematicity here concerns with whether or not the procedures of the assessment is organized in orderly manner.
 - Why do you think it is systematic / unsystematic?
 - Can you give examples for ... point?

9. In your view, how was the level of objectivity and subjectivity of the assessment? Objectivity and subjectivity in this case mean whether or not the assessment produces objective or subjective results.
 - Why do you think it is objective / subjective?
 - Can you give examples for ... point?
10. In your view, how was the level of efficiency of meaning-oriented assessment? Efficiency here means whether or not the assessment is time-efficient or cost-efficient.
 - Why do you think it is efficient / inefficient?
 - Can you give examples for ... point?
11. In your view, how was the level of detail of meaning-oriented assessment?
 - Why do you think it is detailed / not detailed?
 - Can you give examples for ... point?
12. What do you think about the implementation of meaning-oriented assessment in formative assessment, such as assessing translation assignments and quizzes, in the classroom?
 - Why do you think so?
13. What do you think about the implementation of meaning-oriented assessment for summative assessment, such as assessing midterm or final tests?
 - Why do you think so?
14. Would you use this model of assessment in your own class in the future?
 - Why is it so?
 - Would you rather use other types of assessment? Which assessment?
15. Would you recommend this model assessment to your colleagues or fellow lecturers or institution?
 - Why / Why not?

(turn off the recording device)

Thank you very much. Your participation is much appreciated. I am sure that I have gathered all the data that I need. However, if there is anything I need to clarify further, would it be alright if I send an email to you / schedule another interview with you?

Thank you very much.

APPENDIX 3
WRITTEN CONSENT FORM (for participants)

I, the undersigned, have received explanation from the researcher regarding the background, purpose, significance, and procedures of this research. The researcher has assured that my identity will remain confidential and I can freely withdraw my participation should I choose to. I have been given opportunities to ask questions, and the researcher has responded to them satisfactorily. With this form, I confirm that I am willingly and voluntarily participate in this research with no coercion or undue influences from any parties.

Name of participant

Signature of participant

Name of researcher

Signature of researcher

Date : _____

Place : _____

APPENDIX 4 RESEARCH BACKGROUND (for raters)

Meaning-Oriented Assessment: An Analysis on Inter-Rater Reliability, Strengths, and Weaknesses

Suwarni Wijaya Halim

Overview of Research

This interview is conducted as a part of my master study in the Faculty of English Applied Linguistics at Atma Jaya Catholic University. The topic of this research is related to the notion of translation assessment in translation training. There are three aims of this research. First is to find out the result of assessment which is conducted using meaning-oriented assessment by the raters. Second is to discover the reliability of meaning-oriented assessment. Finally, this research tries to identify the raters' perceptions about meaning-oriented assessment. This interview will provide insights and information to achieve the third aim.

Participation & Confidentiality

Your identity as the rater for this project and as the source for this interview will be treated as confidential, and you are free to withdraw from this project should you wish to. The scoring sheets that you submit at the end of the assessment process will be included into the research report. The interview process will take place for approximately an hour and later on will be recorded using a recording device. The audio recording will later on be transcribed, and the transcription will be sent to you so that you will be able to verify your responses before it is used for the research process. The recording itself will be stored in the place where only the researcher can access, and it will be disposed five years after the researcher finalizes her research process.

Benefits

There is no immediate benefit that you can gain from this research. Nevertheless, this research is expected to provide insights for you regarding an alternative form of translation assessment, which is meaning-oriented assessment. This research provides information on mechanisms and details of conducting meaning-oriented assessment, which you can consider using when assessing students' translation works.

Risks

This research is expected to take copious amount of your time. Other than the allocation of time, there is no other risk involved for participating in this research.

Further Questions

If you require further information about this research, you are welcomed to email the researcher (Suwarni Wijaya Halim) at suwarni.halim@gmail.com.

Thank you very much for your participation in this research.

APPENDIX 5
WRITTEN CONSENT FORM (for raters)

**Meaning-Oriented Assessment: An Analysis on Inter-Rater Reliability,
Strengths, and Weaknesses**

Suwarni Wijaya Halim

Statement of Consent

By signing this form, it proves that:

- You have received explanation from the researcher regarding the background, purpose, significance, and procedures of this research;
- You have understood the details of your participation in this research;
- You are aware of the benefits and risks that you will experience;
- You have been given opportunities to ask questions, and the researcher has responded to them satisfactorily;
- You are aware that you are free to withdraw your participation from this research;
- You are aware that the scoring sheets that you will produce will be included in the researcher's thesis, and you agree to this;
- You are aware that the interview will be recorded, transcribed, and included in the researcher's thesis, and you agree to this;
- You understand that the data you provide for this thesis may be used by other people in the future as comparison for their own research;
- You confirm that you are willingly and voluntarily participate in this research with no coercion or undue influences from any parties.

Name : _____

Signature : _____

APPENDIX 6
SAMPLE OF ASSESSMENT SHEET (Student 1)

| No. | Source Text | Target Text | Error | | | Points | |
|-------|---|--|-------|-------|--------|--------|------|
| | | | | Lexis | Clause | | Text |
| 1 | Don't Let Drugs Rule the Nation | Jangan Biarkan Narkoba Menguasai Negara | Ex | A | | | |
| | | | | N | | | |
| | | | Lo | A | | | |
| | | | | N | | | |
| | | | Int | A | | | |
| | | | | N | | | |
| | | | Text | A | | | |
| N | | | | | | | |
| Minor | | | | | | | |
| 2 | President Joko "Jokowi" Widodo's commitment to the fight against drugs is part of his responsibility to protect the nation. | Komitmen Presiden Jokowi dalam memerangi narkoba adalah bagian dari tanggung jawabnya untuk melindungi negara. | Ex | A | | | |
| | | | | N | | | |
| | | | Lo | A | | | |
| | | | | N | | | |
| | | | Int | A | | | |
| | | | | N | | | |
| | | | Text | A | | | |
| N | | | | | | | |
| Minor | | | | | | | |
| 3 | Last year he repeatedly warned of "a drug emergency", saying 50 | Tahun lalu ia kembali memperingatkan "darurat narkoba", menegaskan bahwa 50 orang meninggal setiap harinya akibat dari | Ex | A | | | |
| | | | | N | | | |
| | | | Lo | A | | | |

| | | | | | | | | |
|---|--|--|-------|---|--|--|--|--|
| | people died in relation to drug use every day and 4.5 million people were addicted, giving justification to execute drug convicts. | penggunaan narkoba dan 4,5 juta orang telah terikat narkoba, yang berakhir dengan mengeksekusi para pengguna narkoba. | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 4 | At the opening of a Cabinet meeting on Wednesday, the President said he wanted “more intensive, braver, crazier and more comprehensive integrated efforts to eradicate drugs”. | Dalam pembukaan pertemuan Kabinet pada hari Rabu, Pak Presiden mengatakan bahwa ia ingin “tindakan yang lebih intensif, lebih berani, lebih serius, dan lebih terintegrasi dalam mengatasi narkoba.” | | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 5 | During the meeting, which came on the heels of a series of raids that led to the arrest of military and police personnel and, reportedly, a House of Representatives lawmaker, Jokowi declared a war on drugs. | Selama pertemuan tersebut, yang juga membahas tentang keterlibatan penggunaan narkoba diantara personil TNI dan Polri, dan bahkan anggota DPR, Jokowi mendeklarasikan perang terhadap narkoba. | | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 6 | Whether the braver and crazier measures will include | Bukti dari keberanian dan keseriusan Pak Jokowi akan diuji dalam pengesahan hukuman | | A | | | | |
| | | | | N | | | | |

| | | | | | | | | |
|---|--|---|-------|---|--|--|--|--|
| | another round of executions of drug convicts remains to be seen. | mati bagi para pengguna narkoba. | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 7 | But the way the government and law enforcers have combatted drugs so far has not shifted from the textbook: Raids are conducted here and there, most of the time with much fanfare, dozens are arrested and it is business as usual again until the next cycle of raids. | Namun, cara pemerintah dan badan hukum dalam memerangi narkoba sejauh ini tidaklah maksimal: Penangkapan dilakukan disini dan disana, dan membutuhkan biaya yang besar, sangat banyak yang ditangkap dan ini adalah bisnis sampai siklus penangkapan selanjutnya. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 8 | Worse, law enforcement against drug crimes is limited to small fry and is unable to catch the big fish, who use their money or influence to evade justice. | Lebih buruknya, badan hukum yang memerangi narkoba terbatas dalam kasus yang kecil dan tidak mampu dalam menangani kasus besar / penjahat kelas kakap, yang memakai uang mereka untuk memanipulasi hukum. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 9 | Indeed the drug business is so | Nyatanya, bisnis narkoba sangatlah bersifat | Ex | A | | | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|--|--|----|---|--|--|--|----|---|--|--|----|---|--|--|--|-----|---|--|--|-----|---|--|--|--|------|---|--|--|------|---|--|--|--|-------|--|--|--|-------|--|--|--|--|
| | secretive that law enforcers may find it difficult to capture the kingpins, but after so many raids and tons of drugs have been discovered, the public has yet to hear of any drug ring leader being arrested and prosecuted. | rahasia sehingga badan hukum kesulitan untuk menangkap pemimpinnya, tetapi setelah banyaknya penangkapan dan penemuan bukti narkoba, masyarakat belum juga mendengar kabar penangkapan pemimpin bisnis narkoba tersebut. | <table border="1"> <tr> <td></td> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Lo</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Int</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Text</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Minor</td> <td></td> <td></td> <td></td> <td></td> </tr> </table> | | N | | | | Lo | A | | | | N | | | | Int | A | | | | N | | | | Text | A | | | | N | | | | Minor | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lo | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Int | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Text | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Minor | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10 | Challenges facing drug eradication also include impunity or at least light punishments. | Tantangan dalam memberantas narkoba juga termasuk adanya pembebasan atau setidaknya hukuman ringan. | <table border="1"> <tr> <td rowspan="2">Ex</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Lo</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Int</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Text</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Minor</td> <td></td> <td></td> <td></td> <td></td> </tr> </table> | Ex | A | | | | N | | | | Lo | A | | | | N | | | | Int | A | | | | N | | | | Text | A | | | | N | | | | Minor | | | | |
| Ex | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lo | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Int | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Text | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Minor | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 11 | Many would remember an Army officer, a general's son, who was found guilty of drug abuse more than 15 years ago but had his prison sentence reduced and his dismissal from the military revoked. | Banyak orang akan ingat kasus seorang tentara, seorang anak Jendral, yang terbukti bersalah dalam kasus penggunaan narkoba selama lebih dari 15 tahun, tetapi masa penahanannya dikurangi dan ia kembali diterima dalam militer sebagai tentara. | <table border="1"> <tr> <td rowspan="2">Ex</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Lo</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Int</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td rowspan="2">Text</td> <td>A</td> <td></td> <td></td> <td></td> </tr> <tr> <td>N</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Minor</td> <td></td> <td></td> <td></td> <td></td> </tr> </table> | Ex | A | | | | N | | | | Lo | A | | | | N | | | | Int | A | | | | N | | | | Text | A | | | | N | | | | Minor | | | | |
| Ex | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lo | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Int | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Text | A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Minor | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | |
|----|---|--|-------|---|--|--|--|--|
| 12 | He has won promotions since then. | Ia bahkan dipromosikan sejak saat itu. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 13 | Abuse of power is another cause for concern, especially in the antidrug campaign. | Penyalahgunaan kekuasaan adalah salah satu keraguan tersebut, terutama dalam kampanye anti narkoba. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| | | | Minor | | | | | |
| 14 | This ranges from reports of people without any history of drug use being caught up in drug checks and forced to pay huge sums of money for their release, as well as law enforcers who act as drug dealers or protect drug rings. | Ini diketahui dari laporan orang-orang yang tidak memiliki rekam jejak penggunaan narkoba tetapi ditangkap dan dipaksa untuk membayar mahal supaya mereka bisa bebas dari hukuman, seolah-olah badan hukum yang bertindak sebagai penjual narkoba atau melindungi bisnis narkoba tersebut. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |

| | | | | | | | | |
|------|--|---|-------|---|--|--|--|--|
| | | | Minor | | | | | |
| 15 | Now that the President has called for integrated action to win the war on drugs, the National Narcotics Agency needs to draw up a clear strategy and set concrete targets. | Saat ini, sejak Pak Presiden telah menuntut tindakan yang terintegrasi untuk memenangi perang melawan narkoba, Badan Narkotika Nasional (BNN) membutuhkan strategi yang jelas dan menetapkan target yang konkrit. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| Text | A | | | | | | | |
| | N | | | | | | | |
| | | | Minor | | | | | |
| 16 | The agency must have identified hotspots where drugs are trafficked and circulated, as well as the people who are involved in the drug supply chain. | Agen BNN harus mengidentifikasi tempat / lokasi dimana narkoba diselundupkan dan diperjualbelikan, dan juga pihak-pihak yang terlibat dalam pengadaan narkoba tersebut. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| Text | A | | | | | | | |
| | N | | | | | | | |
| | | | Minor | | | | | |
| 17 | Tightening supervision of prisons, which the President believes are centers of drug trafficking, is mandatory. | Mengetatkan aturan penahanan, yang Pak Presiden percayai adalah sumber dari penjualan narkoba, adalah suatu keharusan. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| Text | A | | | | | | | |
| | N | | | | | | | |

| | | | | | | | | |
|-------|---|--|-------|---|--|--|--|--|
| | | | | N | | | | |
| | | | Minor | | | | | |
| 18 | Random checks are a must — every day if possible, rather than every month as suggested by the President — given the scale of drug abuse facing the country. | Pengecekan adalah suatu keharusan—setiap hari jika diperlukan, lebih baik daripada setiap bulan seperti yang disarankan Pak Presiden— dan dapat memberikan perlawanan terhadap penyalahgunaan narkoba. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| Minor | | | | | | | | |
| 19 | The magnitude of the national campaign against drugs may require a powerful agency like the US Drug Enforcement Administration. | Gelombang kampanye nasional terhadap narkoba membutuhkan agen yang kuat seperti Drug Enforcement Administration milik Amerika Serikat. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |
| | | | Text | A | | | | |
| | | | | N | | | | |
| Minor | | | | | | | | |
| 20 | A debate over this possibility should start sooner rather than later, simply because the war on drugs has already begun. | Perdebatan atas segala kemungkinan ini seharusnya dimulai secepatnya, karena perang terhadap narkoba sudah dimulai. | Ex | A | | | | |
| | | | | N | | | | |
| | | | Lo | A | | | | |
| | | | | N | | | | |
| | | | Int | A | | | | |
| | | | | N | | | | |

APPENDIX 7 GUIDE TO MEANING-ORIENTED ASSESSMENT

A. Overview

Meaning-oriented assessment is a model of translation assessment proposed by Kim (2009), and it is based on Systemic Functional Linguistics (SFL) by Halliday. In assessing translation works, three elements of SFL are emphasized: **Register**, **Metafunction**, and **Lexicogrammatical Elements**. Each element is related to each other, and the relation is described in the chart below.

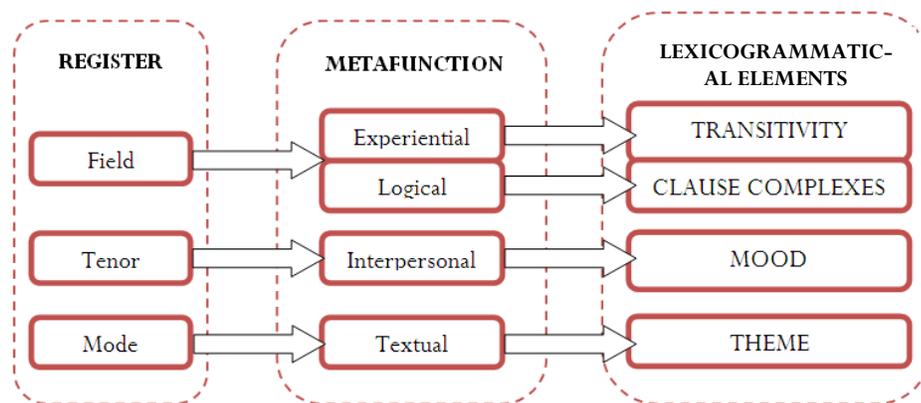


Chart 1. Relation among Register, Metafunction, and Lexicogrammatical Elements (adapted from Kim (2009))

B. Using Meaning-oriented Assessment

Register reflects the context of the text, and it is divided into three: Field, Tenor and Mode. According to Halliday and Matthiessen (2014), **Field** is the topic / focus of the text (i.e. what is being discussed in the text). **Tenor** is relation between the interactants (i.e. social distance, familiarity, and status between the writer and readers). **Mode** is medium of communication (i.e. written or spoken or combination).

In the translation project held by the researcher, Register is reflected through the **translation brief** given to the students during the translation process. The students were required to translate an editorial text taken from *the Jakarta Post*, entitled “Don’t Let Drugs Rule the Nation.” The text below is the content of translation brief.

“The source text (ST) is an editorial taken from the online portal of *the Jakarta Post*. The text tells about the efforts of the government to fight drugs. The main function of the ST is to give information for the newspaper readers, who have sufficient English proficiency and knowledge of the topic. The ST should be translated into Indonesian. The target text (TT) would be published in the online portal of other newspaper, which means that the TT should follow the formal tone of the source text. The function of the TT should follow the ST, which is informative, so the content of the TT must remain

faithful to the ST. The TT would be read by newspaper readers whose mother tongue is Indonesian and who have enough knowledge about the topic.”

Based on the brief, it can be seen that the Field of the ST is related to the attempts by the government to combat drugs in Indonesia. The Tenor between the ST writer and ST readers is formal, in which the writer acts as an expert, and the readers can be considered as knowledgeable of the topic. The Mode of the ST is written and monologic since the text is an article in the newspaper portal. The brief also demands the translations to be faithful to the source text.

Raters should : ensure that the Register of the students’ translation reflects the Register in ST, which is expressed through the translation brief. The raters need to check whether the Field, Tenor, and Mode of the TT correspond with the Field, Tenor, and Mode of the ST.

C. Criteria in Meaning-oriented Assessment

Marking scheme used in meaning-oriented assessment is constructed from the combination of SFL theories and NAATI scoring system.

| | | | Lexis | Clause | Text |
|-------|---|-------------|---------|---------|---------|
| Major | Experiential | Accuracy | 1-2 pts | 2-3 pts | |
| | | Naturalness | 1-2 pts | 2-3 pts | |
| | Logical | Accuracy | | 1-3 pts | |
| | | Naturalness | | 1-3 pts | |
| | Interpersonal | Accuracy | 1-2 pts | | 3-5 pts |
| | | Naturalness | 1-2 pts | | 3-5 pts |
| | Textual | Accuracy | | 1-2 pts | 3-5 pts |
| | | Naturalness | | 1-2 pts | 3-5 pts |
| Minor | Graphological mistakes such as spelling | | | | 0.5 |
| | Minor grammar mistakes that do not impact meaning | | | | 0.5 |

Table 1. Criteria for Meaning-oriented Assessment (taken from Kim (2009, p. 136))

Meaning-oriented assessment uses the system of **point reduction**. The maximum score is **45 points**. If there are any errors found in the translation, the points will be deducted based on the type of errors specified above.

Errors in meaning-oriented assessment are classified into two types: major and minor. **Major errors** deal with deviation in metafunctions, which results in wrong or incorrect transfer of meaning. These major errors will be analyzed further to find out whether they are problematic in terms of **accuracy** or **naturalness**. According to Larson (1984), a translation is **accurate** when the meaning from the source text is exactly conveyed without adding, omitting, or changing any sense in the meaning. On the other hand, a translation can be considered as **natural** if the translation uses

grammatical forms or construction and word pairs or collocation that are adequate and acceptable in target language.

After deciding whether the error is a case of inaccuracy or unnaturalness, the error needs to be analyzed more closely in order to detect the impact of the error. As seen in Table 1, the errors might impact the level of **lexis** (words), **clause**, or **text**. If the error is limited to lexical level and does not affect the clause or text, the range of point reduction below **lexis** column should be used. If the error influences the whole clause but not the whole text, the range of point reduction below **clause** column should be utilized. If the error affects the whole text, the range of point reduction below **text** column should be used.

Minor errors, on the other hand, deal with small mistakes such as misspellings, typos, or grammatical mistakes. These types of errors do not affect the transfer of meanings. The example below is an example of Minor errors.

| | |
|-------------|---|
| Source Text | <i>President</i> Joko “Jokowi” Widodo <i>commitment</i> to fight against drugs is part of his responsibility to protect the nation. |
| Target Text | <i>Komitment</i> <i>President</i> Joko “Jokowi” Widodo dalam memerangi narkoba merupakan bagian dari tanggung jawab beliau dalam melindungi bangsa. |

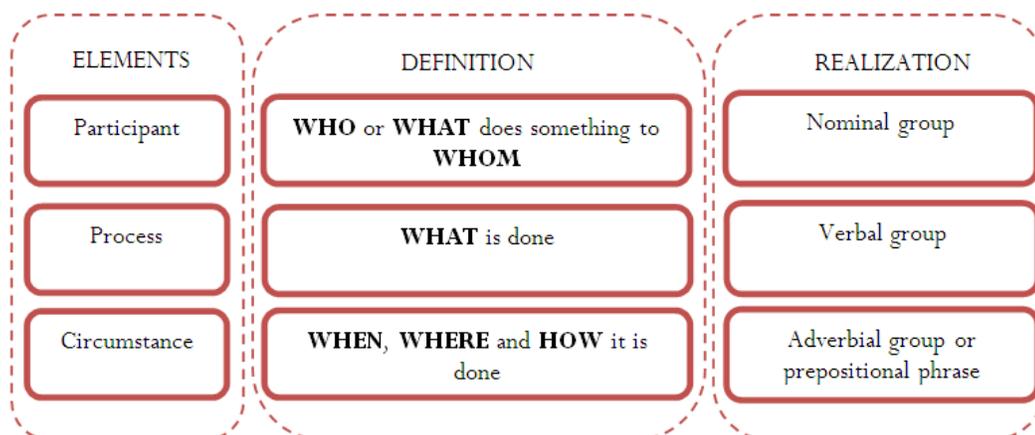
As seen above, the student misspelled the words *komitmen* and *presiden* into *komitment* and *president*. These mistakes, however, do not hamper the transfer of meaning, and readers would most probably still understand the content of the message despite the mistakes. Therefore, 0.5 point will be reduced for *komitment*, and another 0.5 point will be reduced for *president*.

Raters should : remember that they do not merely focus on grammatical mistakes. The most important point is the accuracy and the naturalness of the translation and the accordance of the context of the translation to the context of the source text. Errors in naturalness will constitute smaller point reduction compared to errors in accuracy.

Raters should : also remember that in the process of translating, some changes or adjustments are bound to happen due to the fact that languages are different from one another. Changes such as omission or addition are considered as translation strategies; therefore, if they do not compromise the accuracy and naturalness of the meaning transfer, they should not be considered as an error.

Experiential Metafunction Errors

Experiential metafunction deals with the construction of the reality in the text, and it is reflected through the aspect of **TRANSITIVITY**. Experiential metafunction can be divided into three elements, which is described in the chart below.



| | | | | |
|---------------|-------------|---------------|-----------------------|----------------|
| My sister | is planting | a maple tree | in front of her house | at the moment. |
| Participant 1 | Process | Participant 2 | Circumstance 1 | Circumstance 2 |

Chart 2. Participant, Process, and Circumstances (adapted from Halliday & Matthiessen, 2014)

Despite the fact that SFL entails more detailed analysis on types of Participant, Process and Circumstance, Kim (2009; 2010) focuses only on the classification of transitivity elements and the translation of those elements from source language to target language.

Raters should : break down clause complexes (constituents that consist more than one clause) into several sub-clauses before assigning the elements of Participant, Process, and Circumstance to the constituents. For example, the clause complex “Tightening supervision of prisons, which the President believes are centers of drug trafficking, is mandatory” is composed of two sub-clauses: (1) *tightening supervision of prisons is mandatory*; and (2) *which the President believes are centers of drug trafficking*.

After breaking down the clauses in ST and TT into those constituents, the raters have to analyze whether or not any constituents are translated inaccurately or unnaturally. Afterwards, those experiential errors are analyzed further to determine the effect, whether they might affect the transfer of meaning in the level of lexis or the level of clause. Then the raters can deduct the points based on the information specified in the marking scheme below.

| | | | Lexis | Clause | Text |
|---------|---|-------------|---------|---------|---------|
| Major | Experiential | Accuracy | 1-2 pts | 2-3 pts | |
| | | Naturalness | 1-2 pts | 2-3 pts | |
| | Logical | Accuracy | | 1-3 pts | |
| | | Naturalness | | 1-3 pts | |
| | Interpersonal | Accuracy | 1-2 pts | | 3-5 pts |
| | | Naturalness | 1-2 pts | | 3-5 pts |
| Textual | Accuracy | | 1-2 pts | 3-5 pts | |
| | Naturalness | | 1-2 pts | 3-5 pts | |
| Minor | Graphological mistakes such as spelling | | | | 0.5 |
| | Minor grammar mistakes that do not impact meaning | | | | 0.5 |

Table 2. Criteria for Meaning-oriented Assessment (taken from Kim (2009, p. 136))

Examples of Experientially Inaccurate Translation

1. Error in the level of lexis

This section provides examples of experientially inaccurate translation. Please have a look at the sample below. The experiential error is indicated with italic.

| | | | |
|-----------------------|--|---------------|---|
| Source Text | Challenge facing drug eradication | also include | <i>impunity</i> or at least light punishments. |
| | Participant 1 | Process | Participant 2 |
| Erroneous Translation | Rintangan dalam menghadapi pemberantasan narkoba | juga termasuk | <i>kecacatan hukum</i> atau setidaknya hukuman yang ringan. |
| | Participant 1 | Process | Participant 2 |
| Suggested Translation | Rintangan dalam menghadapi pemberantasan narkoba | juga termasuk | <i>pembebasan dari hukuman</i> atau setidaknya hukuman yang ringan. |
| | Participant 1 | Process | Participant 2 |

After both ST and TT are broken down into clauses, the raters need to compare the ST and TT to identify which component is erroneous. In the case above, the word *impunity* is inaccurately translated into *kecacatan hukum*. According to Merriam Webster online dictionary, *impunity* means “exemption or freedom from punishment, harm, or loss,” which essentially has different meaning from *kecacatan hukum*. The translation did not render the accurate meaning of the word *impunity*. Consequently, this error can be categorized as an error in accuracy, and it happens in the level of lexis and does not affect the whole clause. Since the error does not really affect the conveyance of the whole clause, based on the marking scheme, 2 point should be reduced.

2. Error in the level of clause

Below is an example of experientially inaccurate translation in the level of clause.

| | | | |
|-----------------------|----------------|--------------------------------------|---|
| Source Text | ... the public | <i>has yet to hear of</i> | any drug ring leader being arrested and prosecuted. |
| | Participant 1 | Process | Participant 2 |
| Erroneous Translation | ... masyarakat | <i>telah mendengar mengenai</i> | penangkapan dan penuntutan bandar narkoba. |
| | Participant 1 | Process | Participant 2 |
| Suggested Translation | ... masyarakat | <i>belum juga mendengar mengenai</i> | penangkapan dan penuntutan bandar narkoba manapun. |
| | Participant 1 | Process | Participant 2 |

From the table above, it can be seen that the student inaccurately rendered the process *has yet to hear* into *telah mendengar*. According to Macmillan Dictionary, the phrase *has yet to* is used to indicate that “something has not happened or been done up to the present time, especially when you think it should have happened or been done.” As a result, the opposite of *telah mendengar* is the most appropriate rendition of the process *has yet to hear*, which is *belum juga mendengar*. The student has inaccurately translated the element of process, and this error can affect the meaning of the whole clause. Consequently, this error deserves the reduction of 3 points since it compromises the delivery of meaning of the whole clause.

Examples of Experientially Unnatural Translation

1. Error in the level of lexis

This section provides examples of experientially unnatural translation. Please have a look at the sample below. The experiential error is indicated with italic.

| | | | |
|-----------------------|---------------|--------------|--|
| Source Text | ... he | wanted | “more intensive, braver, crazier and more comprehensive integrated efforts to <i>eradicate drugs</i> ” |
| | Participant 1 | Process | Participant 2 |
| Erroneous Translation | ... beliau | menginginkan | “usaha terpadu yang lebih intensif, berani, gila, dan komprehensif dalam <i>membasmi narkoba</i> ” |
| | Participant 1 | Process | Participant 2 |
| Suggested Translation | ... beliau | menginginkan | “usaha terpadu yang lebih intensif, berani, gila, dan komprehensif dalam |

| | | | |
|--|---------------|---------|------------------------------|
| | | | <i>memberantas narkoba</i> ” |
| | Participant 1 | Process | Participant 2 |

As seen in the table above, the translator translated *eradicate drugs* into *membasmi narkoba*. According to Merriam Webster Online Dictionary, *to eradicate* means *to destroy* or *to eliminate*. The translation *membasmi* essentially has the same sense with the source text *eradicate*. However, the combination of *membasmi* and *narkoba* is not common in Bahasa Indonesia, so the translation sounds unnatural. This error in naturalness only happens in the lexical level, and it does not affect the transfer of message in the level of clause. Thus, based on the marking scheme and based on the understanding that error in naturalness has smaller point reduction compared to error in accuracy, 1 point should be reduced.

2. Error in the level of clause

Below is an example of experientially unnatural translation in the level of clause.

| | | | | |
|-----------------------|---------------|--------------------------|-----------------------------|-----------------|
| Source Text | He | <i>has won</i> | <i>promotions</i> | since then. |
| | Participant 1 | Process | Participant 2 | Circumstance |
| Erroneous Translation | Dia | <i>telah memenangkan</i> | <i>pengangkatan jabatan</i> | sejak saat itu. |
| | Participant 1 | Process | Participant 2 | Circumstance |
| Suggested Translation | Dia | <i>telah naik</i> | <i>pangkat</i> | sejak saat itu. |
| | Participant 1 | Process | Participant 2 | Circumstance |

As seen in the table above, the student translated *has won promotions* into *telah memenangkan*. Although the student tried to render the process and participant into the closest equivalent in Indonesian, the translation was conducted too literally, and it results in awkward and unnatural translation. The suggested alternative would be *telah naik pangkat* which is more common in Indonesian. Despite the fact that the translation is unnatural, the readers most probably would be able to understand the translation. It, however, affects the level larger than lexis, which is the clause. Based on the marking scheme and based on the understanding that error in naturalness has smaller point reduction compared to error in accuracy, 2 point should be reduced.

Logical Metafunction Error

Logical metafunction is related to experiential metafunction, and it deals with **logical links** among **clause complexes** (cluster of two or more clauses). Clauses which are

constructed of single clause unit do not need to be analyzed for the logical metafunction. Unlike experiential analysis, the raters do not need to classify the components of the clause. The analysis on logical metafunction can be divided into two kinds: **taxis relation** and **logico-semantic relation**.

Taxis relation deals with the structural organization of the clause complexes. It can be divided further into two types: **parataxis** and **hypotaxis**. **Parataxis** is characterized through the existence of more than one independent clause combined together through the usage of coordinating conjunctions (*for, and, nor, but, or, yet, so, neither ... nor, and either ... or*) and certain punctuations (comma, colon, and semi-colon). **Hypotaxis**, on the other hand, is signaled through the existence of one independent clause and more than one dependent clause, which can be identified through the usage of subordinating conjunctions, relative pronoun, and non-finite clauses (Eggins, 2004).

Logico-semantic relation deals with the semantic organization of the clause complexes. It can also be divided further into two types: **projection** and **expansion**. **Projection** is “the logico-semantics of quoting and reporting speech and thoughts” whereas **expansion** is “the logico-semantics of developing on previous meanings” (Eggins, 2004, p. 271). The relation of **projecting** can be expressed through the verbs of “saying or thinking (or any of their many synonyms)” (Eggins, 2004, p. 271). If such verbs are not present in the clause complex, there is big possibility that the logico-semantic relation is expansion. According to Eggins (2004, pp. 279-284), **expansion** can be divided further into three types:

- **Elaboration** → “one clause elaborates on the meaning of another by restating the meaning in different words, developing the meaning to become more specific or clarifying by giving explanation”
- **Extension** → “one clause extends the meaning of another by adding something new”
- **Enhancing** → “one clause enhances the meaning of another by utilizing reference to time, space, manner, cause or condition”

Raters should : (1) identify the clause complexes in the source text; (2) identify the types of taxis, whether the clause complex has parataxis or hypotaxis relation; (3) identify the logico-semantic relation of the clause complex; (4) analyze whether the logico-semantic relation is conveyed or rendered appropriately in the target text.

As stated in the marking scheme below, Logical metafunction operates in the level of clause, and the range of point reduction goes from 1 to 3 points.

| | | | Lexis | Clause | Text |
|---------|---|-------------|---------|---------|---------|
| Major | Experiential | Accuracy | 1-2 pts | 2-3 pts | |
| | | Naturalness | 1-2 pts | 2-3 pts | |
| | Logical | Accuracy | | 1-3 pts | |
| | | Naturalness | | 1-3 pts | |
| | Interpersonal | Accuracy | 1-2 pts | | 3-5 pts |
| | | Naturalness | 1-2 pts | | 3-5 pts |
| Textual | Accuracy | | 1-2 pts | 3-5 pts | |
| | Naturalness | | 1-2 pts | 3-5 pts | |
| Minor | Graphological mistakes such as spelling | | | | 0.5 |
| | Minor grammar mistakes that do not impact meaning | | | | 0.5 |

Table 3. Criteria for Meaning-oriented Assessment (taken from Kim (2009, p. 136))

Examples of Logically Inaccurate Translation

This section provides examples of logically inaccurate translation. Please have a look at the sample below. The logical error is indicated with italic.

| | |
|-----------------------|---|
| Source Text | Last year he repeatedly warned of “a drug emergency”, <i>saying</i> 50 people died in relation to drug use every day ... |
| Erroneous Translation | Tahun lalu, beliau berkali-kali memperingatkan mengenai “Bahaya Narkotika” <i>yang menyatakan</i> bahwa 50 orang meninggal setiap hari terkait penggunaan obat-obatan terlarang ... |
| Suggested Translation | Tahun lalu, beliau berkali-kali memperingatkan mengenai “Bahaya Narkotika” <i>dengan menyatakan</i> bahwa 50 orang meninggal setiap hari terkait penggunaan obat-obatan terlarang ... |

The source text indicated in the table above is a case of clause complex with one independent and one dependent (specifically, non-finite) clause. According to Kim (2009, p. 139), the logico-semantic relation in non-finite clauses “may not always be clear.” In the source text, the dependent clause starts with *saying*, which is a verb that reflects projection. It can be seen also that the student managed to translate *saying* with verb *menyatakan*, which is also a verb of projection in Indonesian.

However, the problem actually occurs because the student added *yang* in the translation, which shifted the logico-semantic relation from projection into expansion. The verb *saying* is supposed to project the statement from “he.” By translating it into *yang menyatakan*, the translation instead expands the meaning of “a drug emergency” instead of “he.” In other words, the non-finite is translated inaccurately, and it changes the relation between the first and second clause. It is therefore treated as a serious error, which results in 3-point reduction.

Examples of Logically Unnatural Translation

This section provides examples of logically unnatural translation. Please have a look at the sample below. The logical error is indicated with italic.

| | |
|-----------------------|--|
| Source Text | Last year he repeatedly warned of “a drug emergency”, saying 50 people died in relation to drug use every day, and 4.5 million people were addicted, <i>giving</i> justification to execute drug convicts. |
| Erroneous Text | Tahun lalu, beliau berulang kali memperingatkan mengenai “Bahaya Narkotika” dengan menyatakan bahwa 50 orang meninggal setiap hari terkait dengan penggunaan narkoba, dan 4.5 juta orang mengalami kecanduan narkoba, <i>memberikan</i> justifikasi untuk menghukum narapidana kasus narkoba. |
| Suggested Translation | Tahun lalu, beliau berulang kali memperingatkan mengenai “Bahaya Narkotika” dengan menyatakan bahwa 50 orang meninggal setiap hari terkait dengan penggunaan narkoba, dan 4.5 juta orang mengalami kecanduan narkoba. <i>Fakta-fakta tersebut memberikan</i> justifikasi untuk menghukum narapidana kasus narkoba. |

As seen in the source text, it is actually quite frequent in English to use several non-finite clauses in one construction. However, that form of construction is rarely used in Indonesian, and keeping the translation that way might hamper the readers’ comprehension of the target text. Furthermore, the translation would sound unnatural since Indonesian people seldom use this kind of construction. The most appropriate step is to conduct a translation shift by dividing the clause from the independent clause and forming a new clause. The student, however, maintained the non-finite structure and risked the natural delivery of the message and comprehension of the target readers. It is therefore treated as an error, which results in 1-point reduction from the total 45 points.

Interpersonal Metafunction Error

Interpersonal metafunction deals with the elements of interaction within the text. It is reflected through the system of **MOOD**, which can be classified further into **Mood** and **modality**. Mood is related to the structure of the clauses. There are three types of Mood: **declarative**, **imperative**, and **interrogative** (Martin, Matthiessen & Painter, 2010). The text used in this research used declarative Mood in the overall text, and the students also translated the whole clauses in the text in the form of declarative Mood. This element of interpersonal metafunction, therefore, does not need to be observed.

Modality, on the other hand, involves the usage of **modal finite** (e.g. he *might* go, she *would* leave) and **adjuncts** in the form of adverbials (e.g. she *rarely* ate, they have *probably* left the house) in the clause. Modality in SFL is divided into two

types: **modalization** (which includes modals of probability and usuality) and **modulation** (which includes modals of obligation and inclination).

Modals of probability describe the likelihood of something to happen while the **modals of usuality** indicate the frequency of something to happen. **Modals of obligation** express the necessity of doing something whereas **modals of inclination** provide information about someone’s willingness to do something (Halliday & Matthiessen, 2014). Below is the diagram of modality which specifies the varieties of functions that modals can indicate.

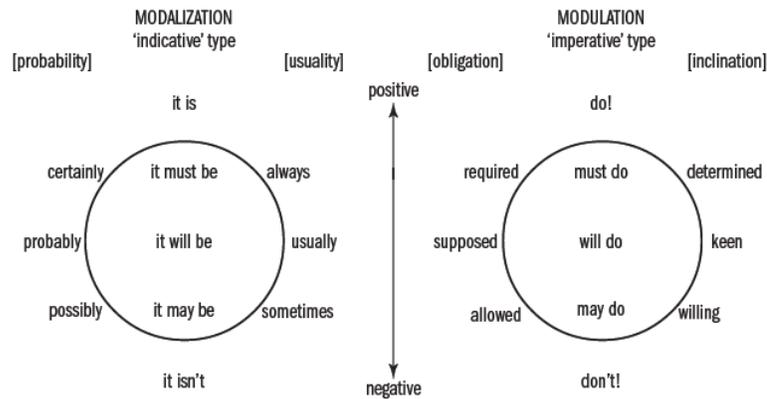


Chart 3. Modalization and Modulation (taken from Halliday & Matthiessen, 2014, p. 691)

The arrow in the middle of two circles indicates the degree. The lower the position of the modals or adjuncts in the circle, the lower (more negative) the probability, usuality, obligation, and inclination are. In contrast, the higher the position of the modals or adjuncts in the circle, the higher (more positive) the probability, usuality, obligation, and inclination are. The table below further elaborates the types of modal finite and their position.

| | Low | Median | High |
|-----------------|---|--|---|
| Positive | can, may, could, might, (dare) | will, would, should, is / was to | must, ought to, need, has / had to |
| Negative | needn't, doesn't / didn't + need to / have to | won't, wouldn't, shouldn't (isn't / wasn't to) | mustn't, oughtn't to, can't, couldn't, (mayn't, mightn't, hasn't / hadn't to) |

Table 4. Modal Finite Operators (taken from Halliday & Matthiessen, 2014, p. 145)

Modals in English are usually translated into verbs or adverbials in Indonesian because Indonesian does not have the system of modalization (Alwi, 1992). The difference in terms of parts of speech, therefore, should not be an issue since English and Indonesian have different grammatical systems.

Raters should : identify clauses in the text that contains modalization (both modal finite and mood adjunct). After the modalization is found, the raters should identify the position and degree of modalization in Chart 3 and Table 4. Then, the raters need to analyze the translation in order to see whether the degree of the translation of modalization matches the source text.

Besides modality, interpersonal metafunction also deals with the Tenor of the text, which includes **politeness, formality scales, social distance, and power relation**. Those elements can be inferred from the situational and cultural context of the text, in this case, through the observation towards the **role and status** of the interactants, i.e. the author and the readers of the text. When one interactant has higher status and power than the other interactant or when both interactants are not very familiar with each other, the language used when they communicate and address each other would differ greatly in terms of politeness.

Raters should : analyze the Tenor of the text carefully in order to identify the nature of relationship between the author and the readers. This can be seen from how the author addresses the readers and communicate with them in the text. Furthermore, the raters need to see whether the students transfer the same nature of the relationship from the source text into the target text.

| | | | Lexis | Clause | Text |
|---------|---|-------------|---------|---------|---------|
| Major | Experiential | Accuracy | 1-2 pts | 2-3 pts | |
| | | Naturalness | 1-2 pts | 2-3 pts | |
| | Logical | Accuracy | | 1-3 pts | |
| | | Naturalness | | 1-3 pts | |
| | Interpersonal | Accuracy | 1-2 pts | | 3-5 pts |
| | | Naturalness | 1-2 pts | | 3-5 pts |
| Textual | Accuracy | | 1-2 pts | 3-5 pts | |
| | Naturalness | | 1-2 pts | 3-5 pts | |
| Minor | Graphological mistakes such as spelling | | | | 0.5 |
| | Minor grammar mistakes that do not impact meaning | | | | 0.5 |

Table 5. Criteria for Meaning-oriented Assessment (taken from Kim (2009, p. 136))

Based on the marking scheme, it can be seen that errors in Interpersonal metafunction might affect the level of lexis or the level of text. Errors in level of lexis might cause 1- to 2-point reduction while errors in the level of text might cause 3- to 5-point reduction.

Examples of Interpersonally Inaccurate Translation

1. Error in the level of lexis

This section provides examples of interpersonally inaccurate translation. Please have a look at the sample taken from paragraph 1 below. The interpersonal error is indicated with italic.

| | |
|-----------------------|--|
| Source Text | Indeed the drug business is so secretive that law enforcers <i>may</i> find it difficult to capture the kingpins... |
| Erroneous Translation | Bisnis narkoba memang sangat rahasia sehingga aparat penegak hukum <i>akan</i> mengalami kesulitan untuk menangkap para gembong narkoba... |
| Suggested Translation | Bisnis narkoba memang sangat rahasia sehingga aparat penegak hukum <i> mungkin</i> mengalami kesulitan untuk menangkap para gembong narkoba... |

The student translated the modal *may* into *akan*. The translation of the modal can be considered as erroneous and inaccurate. Based on the analysis towards the clause, it can be concluded that the usage of modal *may* expresses probability. According to the diagram by Halliday & Matthiessen (2014), modal *may* for probability is placed in the area of Modalization at the lower position, which indicates negative degree of probability. However, the translation *akan* indicates higher degree of probability since it is placed at the middle position. Despite the inaccuracy of translation, the error is only limited to lexical level and does not affect the text in larger scale. Since the translation might cause miscomprehension and misunderstanding of the readers, 2-point reduction is deemed necessary.

2. Error in the level of text

Below is an example of interpersonally inaccurate translation in the level of clause.

| | |
|-----------------------|--|
| Source Text | Indeed the drug business is so secretive that law enforcers <i>may</i> find it difficult to capture the kingpins... |
| Erroneous Translation | Bisnis narkoba memang sangat rahasia sehingga aparat penegak hukum mengalami kesulitan untuk menangkap para gembong narkoba... |
| Suggested Translation | Bisnis narkoba memang sangat rahasia sehingga aparat penegak hukum <i> mungkin</i> mengalami kesulitan untuk menangkap para gembong narkoba... |

The source text used modal finite *may* in order to signal low level of probability. The modal finite *may* shows that the writer does not give full claim to his or her statement but instead only expresses his or her opinion between the positive

polarity (*yes, the law enforcers find it difficult to capture the kingpins*) and negative polarity (*no, the law enforcers do not find it difficult to capture the kingpins*) by using modal finite *may*. However, the student failed to render this. The translation creates an impression that the law enforcers do experience difficulty when tracking the kingpins whereas the source text implies that it is a possibility. This error affects larger area than lexis, and it, therefore, warrants for 3-point deduction.

Examples of Interpersonally Unnatural Translation

1. Error in the level of lexis

This section provides examples of interpersonally unnatural translation. Please have a look at the sample taken from paragraph 1 below. The interpersonal error is indicated with italic.

| | |
|-----------------------|---|
| Source Text | Last year <i>he</i> repeatedly warned of “a drug emergency”, saying 50 people died in relation to drug use every day, and 4.5 million people were addicted, giving justification to execute drug convicts. |
| Erroneous Translation | Tahun lalu <i>dia</i> menghimbau berulang kali tentang “Bahaya Narkotika” dengan menyatakan bahwa 50 orang meninggal disebabkan oleh penggunaan obat-obatan terlarang setiap harinya dan 4.5 juta orang kecanduan. Fakta-fakta tersebut memberikan justifikasi untuk menghukum narapidana kasus narkoba. |
| Suggested Translation | Tahun lalu <i>beliau</i> memperingatkan berulang kali tentang “Bahaya Narkotika” dengan menyatakan bahwa 50 orang meninggal disebabkan oleh penggunaan obat-obatan terlarang setiap harinya dan 4.5 juta orang kecanduan. Fakta-fakta tersebut memberikan justifikasi untuk menghukum narapidana kasus narkoba. |

Interpersonal metafunction reflects the tenor, which deals with the politeness, formality scales, and power relation. The student translated the pronoun *he*, which refers to the president of Indonesia, into *dia*. This translation is considered as unnatural since it neglects the formality scales. Indonesian language has variety of terms to address somebody, depending on the formality scales and the power relation. In this case, since the style of translation has to be formal, and the source author referred to someone with higher power and status, the translation of *he* into *dia* is considered as inappropriate since *dia* is usually used to refer to someone with equal or weaker power. However, this error does not cause misunderstandings, and most readers will probably comprehend who *dia* refers to. Therefore, only 1 point will be deducted for this error.

2. Error in the level of text

The case of interpersonally unnatural errors in the level of text can be treated as an extreme case. If the student translated the whole text without paying attention to the style and formality as suggested in the translation brief (e.g. the whole text is translated in informal style despite the brief specifies formal translation), it can be treated as the problem of unnaturalness in the level of text and warrants for 3 to 5 point reduction depending on the severity of the errors.

Textual Metafunction Error

The lexicogrammatical element of textual metafunction is **Theme**. There are two elements that are involved in the analysis of textual metafunction: **Theme and Rheme** and **Cohesive Devices**.

Theme is the main point or focus of the clause in which the readers focus their attention to. **Rheme**, on the other hand, modifies or gives further information about the Theme. Please have a look at the examples below.

| | |
|--|---|
| For centuries, Yellow canaries Miners In mining To 'test' the air in mining, The air | yellow canaries have been used to 'test' the air in mining. have been used to 'test' the air in mining. have used yellow canaries to 'test' the air for centuries. yellow canaries have been used to 'test' the air for centuries. yellow canaries have been used for centuries. has been 'tested' in mining for centuries by using yellow canaries. |
| Theme | Rheme |

Table 6. Examples of Theme and Rheme (taken from Thompson, 2014, p. 147)

In English, Theme is expressed by the first constituent in the clause. As seen from the examples above, Theme can be in the form of noun phrases, groups of noun phrases, embedded clause, adverbials, prepositional phrases, and complements. The remaining elements are treated as Rheme. Indonesian largely follows the same pattern (Damayanti, 2012; Sukarno, 2014; Sofyan & Rosa, 2014). The constituent in the initial position forms the Theme while the remaining of the clause forms the Rheme.

Raters should : identify the Theme and Rheme of the clause in the source text and check whether the Theme and Rheme are translated accordingly.

The other elements are **cohesive devices**. There are several types of cohesive devices (Halliday & Hasan, 1976). They are:

- **References** occur when one word in a clause refers to another word in the previous or next clause. Examples of references include the usage of definite article (e.g. *the*) and pronouns (e.g. *he, they, we, this, those, there*, etc)
- **Substitution** involves replacing one element in the clause with other elements (e.g. *I have done my homework, and you should do the same.*)
- **Ellipsis** means omission of elements in the clause (e.g. – *Did you lock the door? – I did Ø.*)
- **Conjunctions** and **transition words** are elements that connect one clause with other clauses (e.g. *and, but, which, that, because*, etc) and one paragraph to another (e.g. *however, moreover, then, first*, etc)
- **Lexical cohesion** refers to various process of vocabulary selection. It includes repetition, synonym, and collocation.

Raters should : identify the clauses that contain cohesive devices in the source text. Then, the raters need to check whether those devices are translated accordingly in the target text.

Errors in Textual metafunction may cause problems in the level of clause and text. The raters have to analyze the clause carefully in order to decide whether the errors only compromise the clause or affect the whole text and deduct the points accordingly.

| | | | Lexis | Clause | Text |
|---------|---|-------------|---------|---------|---------|
| Major | Experiential | Accuracy | 1-2 pts | 2-3 pts | |
| | | Naturalness | 1-2 pts | 2-3 pts | |
| | Logical | Accuracy | | 1-3 pts | |
| | | Naturalness | | 1-3 pts | |
| | Interpersonal | Accuracy | 1-2 pts | | 3-5 pts |
| | | Naturalness | 1-2 pts | | 3-5 pts |
| Textual | Accuracy | | 1-2 pts | 3-5 pts | |
| | Naturalness | | 1-2 pts | 3-5 pts | |
| Minor | Graphological mistakes such as spelling | | | | 0.5 |
| | Minor grammar mistakes that do not impact meaning | | | | 0.5 |

Table 7. Criteria for Meaning-oriented Assessment (taken from Kim (2009, p. 136))

Examples of Textually Inaccurate Translation

1. Error in the level of clause

This section provides examples of textually inaccurate translation. Please have a look at the sample below. The interpersonal error is indicated with italic.

| | |
|--------|--|
| Source | Abuse of power is another cause for concern, especially in the |
|--------|--|

| | |
|-----------------------|--|
| Text | antidrug campaign. <i>This</i> ranges from reports of people without any history of drug use being caught up in drug checks... |
| Erroneous Translation | Penyalahgunaan kekuasaan juga salah satu penyebab kekhawatiran, terutama dalam kampanye anti narkoba. <i>Perkumpulan ini</i> bervariasi, dari penangkapan orang-orang yang tidak mempunyai sejarah dalam penggunaan narkoba dalam pemeriksaan narkoba... |
| Suggested Translation | Penyalahgunaan kekuasaan juga salah satu penyebab kekhawatiran, terutama dalam kampanye anti narkoba. <i>Penyalahgunaan tersebut</i> bervariasi, dari penangkapan orang-orang yang tidak mempunyai sejarah dalam penggunaan narkoba dalam pemeriksaan narkoba... |

The problematic part as seen in the table above is the translation of demonstrative reference *this* into *perkumpulan ini*. Demonstrative reference *this* is an element of cohesion which is used to represent other elements such as words, phrases, or clauses. In this case, *this* actually refers to noun phrase *abuse of power* instead of *antidrug campaign*. However, the student misunderstood and translated it into *perkumpulan ini*. This is clearly an error in terms of accuracy. This error does not affect the text as a whole but only affects the level of clause. However, it is severe enough to cause misunderstanding; thus, 2 points are deducted from the total points.

2. Error in the level of text

The case of textually inaccurate errors in the level of text can be treated as an extreme case. If the student translated the whole text without paying attention to the structure of theme and rheme (e.g. rheme is placed at the initial position all throughout the translated text) or translated the whole cohesive devices in the whole text inaccurately (e.g. cohesive device that shows contrast relation in ST is translated as cohesive device that shows addition in TT or the cohesive devices in the whole text are left untranslated), it can be treated as the problem of inaccuracy in the level of text and warrants for 3 to 5 point reduction depending on the severity of the errors.

Examples of Textually Unnatural Translation

1. Error in the level of clause

This section provides examples of textually unnatural translation. Please have a look at the sample below. The interpersonal error is indicated with italic.

| | |
|-------------|---|
| Source Text | Abuse of power is another cause for concern, especially in the antidrug campaign. <i>This</i> ranges from reports of people without any history of drug use being caught up in drug checks... |
|-------------|---|

| | |
|-----------------------|--|
| Erroneous Translation | Penyalahgunaan kekuasaan juga salah satu penyebab kekhawatiran, terutama dalam kampanye anti narkoba. <i>Ini</i> bervariasi, dari penangkapan orang-orang yang tidak mempunyai sejarah dalam penggunaan narkoba dalam pemeriksaan narkoba... |
| Suggested Translation | Penyalahgunaan kekuasaan juga salah satu penyebab kekhawatiran, terutama dalam kampanye anti narkoba. <i>Penyalahgunaan tersebut</i> bervariasi, dari penangkapan orang-orang yang tidak mempunyai sejarah dalam penggunaan narkoba dalam pemeriksaan narkoba... |

In the case above, the problematic part is the translation of *this* into *ini*. The translation is considered unnatural because the translation *ini* does not specify the object it referred to. In Indonesian, it is more preferable to explicate ambiguous elements in order to avoid confusion. In this case, demonstrative pronoun *this* can be ambiguous since it might refer to different elements, and its translation into *ini* does not give sufficient information about the referred elements. This unnatural translation affects the clause; therefore, 1 point should be deducted from the total score.

2. Error in the level of text

Below is another example of textually unnatural translation.

| | |
|-----------------------|--|
| Source Text | Whether the braver and crazier measure will include another round of executions of drug convicts remains to be seen. |
| Erroneous Translation | Apakah tindakan yang lebih berani dan lebih gila tersebut akan mencakup hukuman mati bagi para pengedar narkoba masih belum bisa dipastikan. |
| Suggested Translation | Masih belum bisa dipastikan apakah tindakan yang lebih berani dan lebih gila tersebut akan mencakup hukuman mati bagi para pengedar narkoba. |

In the translation, it can be seen that the student tried to closely follow the pattern of the source text by maintaining the position of Theme (*whether the braver and crazier measure will include another round of executions of drug convicts*) and Rheme (*remains to be seen*) in the translation. This, however, results in an awkward translation. In Indonesian sentential construction, it is rare and unnatural to find noun clauses placed in the initial position. The student's translation is highly marked and might cause confusion among the readers. One way to deal with this is by employing translation shift and adjusting the structure of the clause so that the noun clause is placed in the non-initial position. The student's translation can be considered as unnatural transfer of textual metafunction in the level higher than clause since the translation might affect the

reader's understanding on the text. Therefore, 3 points should be deducted for this error.

References

- A big fish in a small pond. In *Merriam-Webster.com*. Retrieved October 5, 2016, from <http://www.merriamwebster.com/dictionary/a%20big%20fish%20in%20a%20small%20pond>.
- Alwi, H. (1992). *Modalitas dalam bahasa Indonesia* (Modality in bahasa Indonesia) (Doctoral dissertation). Yogyakarta: Kanisius.
- Damayanti, Y. (2012). Theme equivalence and theme shift found in Indonesian-English translation of thesis abstracts. *LANGUAGE CIRCLE Journal of Language and Literature*, 7(1), 33-43.
- Eggs, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). New York, NY: Continuum.
- Eradicate. In *Merriam-Webster.com*. Retrieved October 7, 2016, from <http://www.merriam-webster.com/dictionary/eradicate>.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Oxon, OX: Hodder Arnold.
- Has yet to do something. In *Macmillandictionary.com*. Retrieved October 1, 2016, from <http://www.macmillandictionary.com/dictionary/british/have-yet-to-do-something>.
- Impunity. In *Merriam-Webster.com*. Retrieved October 5, 2016, from <http://www.merriam-webster.com/dictionary/impunity>.
- Kim, M. (2009). Meaning-oriented assessment of translations: SFL and its application to formative assessment. In C. V. Angelelli, & H. E. Jacobson (Eds.), *Testing and assessment in translation and interpreting studies: a call for dialogue between research and practice* (pp. 123-158). Philadelphia, PA: John Benjamins.
- Kim, M. (2010). Translation error analysis: A systemic functional grammar approach, In C. Coffin, T. Lillis, & K. O'Halloran (Eds.), *Applied linguistics methods: A reader* (pp. 84-94). New York, NY: Routledge.
- Larson, M. (1984). *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.
- Martin, J. R., Matthiessen, C. M. I. M., & Painter, C. (2010). *Deploying functional grammar*. Beijing: The Commercial Press.
- Sukarno. (2014). Topic development and thematic patterns for the Indonesian text of Friday sermons. *Humaniora*, 26(3), 315-327.
- Sofyan, R. & Rosa, R. N. (2014). Using theme and rheme analysis in Indonesian-English translation. *International Conference on "Empowering Local Wisdom in Support of National Identity"* (pp. 779-787). Medan: Balai Bahasa of North Sumatera & University of North Sumatera.

The Editorial Board (2016, February 26). Don't Let Drugs Rule the Nation. *The Jakarta Post*. Retrieved from <http://www.thejakartapost.com/news/2016/02/26/editorial-don-t-let-drugs-rule-nation.html>.

Thompson, G. (2014). *Introducing functional grammar* (3rd ed.). New York, NY: Routledge.

**APPENDIX 8
MEETINGS DATABASE**

| Raters | Meeting for Scaffolding | Return of Assessment Sheets | Meeting for Interview | Length of Interview |
|---------------|--------------------------------|------------------------------------|------------------------------|----------------------------|
| Rater 1 | 07 October 2016 | 03 December 2016 | 05 December 2016 | 28 minutes 32 seconds |
| Rater 2 | 27 October 2016 | 30 November 2016 | 13 December 2016 | 52 minutes 21 seconds |
| Rater 3 | 04 November 2016 | 10 January 2017 | 17 January 2017 | 52 minutes 17 seconds |