

2014

# Enhancing rater reliability in the assessment of academic writing: An SFL analysis of raters' rating and perceptions

Gantina Sitepu  
*University of Wollongong*

## **UNIVERSITY OF WOLLONGONG**

### **COPYRIGHT WARNING**

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

# **Enhancing rater reliability in the assessment of academic writing: An SFL analysis of raters' rating and perceptions**

A thesis submitted in fulfilment  
of the requirements of the award of the degree

**Doctor of Education**

from

University of Wollongong

by

**Gantina Sitepu**

S.S. (Medan, Indonesia), M.Ed. (TESOL) (UOW)

Faculty of Social Sciences

2014

## **Certification**

I, Gantina Sitepu, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Education, in the Faculty of Social Sciences, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

---

Gantina Sitepu

27 August 2014

## **Acknowledgements**

My indebtedness, first and foremost, goes to the Endeavour Awards for granting me a scholarship throughout my study. It is my deep honour to be one of the scholarship awardees. My appreciation also extends to the University of Wollongong for providing partial scholarship during the final year of my candidature.

I express my sincere gratitude to my supervisors, Dr Pauline Jones and Associate Professor Honglin Chen, for providing me with inspiration and unwavering support during the long journey of my candidature and throughout the writing of this thesis. I thank them both for their patience and enthusiasm for this research, and for encouraging me at various stages so that I was capable of accomplishing this project. Their constant challenging of my intellectual thinking helped me immeasurably.

I also owe special thanks to all the people involved in the research. I wholeheartedly thank the English for Academic Purposes (EAP) teachers for contributing their valuable time to this study, and the EAP students who kindly gave me permission to use their work. I am additionally grateful to my fellow research students with whom I shared the journey for their friendship and intellectual companionship.

My heartfelt thanks go to my family in Indonesia: my mother, brother and sisters for their sincere love and prayers throughout my candidature. Finally, I dedicate this thesis to my late sister, Sadakata Pandia, with whom I used to share my thoughts.

## **Abstract**

This study investigated and aimed to enhance teachers' assessment practices related to academic writing in a language school in Indonesia. For this purpose, the more explicit and linguistically informed marking criteria were developed and introduced to the teachers' involved in teaching English for Academic Purposes (EAP) classes in the given language school through an intervention. The study adopted an action-oriented case study to capture the teachers' marking experiences, which comprise their accounts of issues of rubrics and marking consistency, and their actual performance in the assessment of argument essays written by 20 EAP students. The study drew on Systemic Functional Linguistics (SFL) and Appraisal theory (Halliday 1978; 1985a, Halliday and Hasan 1985, Halliday and Matthiessen 2004, Martin 1984, Martin and White 2005; 2007) to analyse teachers' accounts and inform the design of the new assessment criteria.

The main purpose of the study is to identify the effect of an SFL informed rubric through training on marking consistency to enhance rater reliability. It aims to design an SFL informed model based on essay analysis and implementing this through the training of teachers. In order to achieve this, specifically, the thesis aims to investigate teachers' pre-perceptions about marking using an existing marking rubric (pre-intervention) and design a new marking model using an SFL framework of genre, Field, Tenor and Mode, and trains the teachers (intervention). Finally, it aims to identify the changes from teachers' pre-perceptions about existing rubrics to their post perceptions about the new rubrics and marking practices.

This study employed a qualitative approach, drawing on six EAP teachers' marking experiences.

The research design comprised three main parts that addressed the research questions. These involved: (i) investigating teachers' perceptions and sources of problems in marking, (ii) designing and implementing the linguistically informed marking criteria, and (iii) investigating teachers' opinions and actual marking practices (pre- and post-intervention).

This study had three major findings. The first was that linguistically informed assessment criteria, which became shared knowledge of teachers, had a key role in enhancing the quality of assessment in academic writing. The improved quality of teachers' assessment was indicated by the higher consistency rates among the teachers in their marking of student essays when using the SFL-informed rubric. The second finding was that teachers response was significantly more positive when talking about their marking experience using the SFL informed rubric than when using the existing rubric. Finally, teachers were well aware of their teaching, marking practices and students' needs, and they recognised the gaps between the three.

The contributions of the study were: (i) it made visible what counts in successful student essays, (ii) it expanded the application of SFL theory by trying it out on the assessment of academic writing in a non-English speaking country, and (iii) it stimulated a discussion on the relationship between teachers' perceptions and marking practices.

# Table of Contents

---

<b>Certification.....</b>	<b>i</b>
<b>Acknowledgements .....</b>	<b>ii</b>
<b>Abstract.....</b>	<b>iii</b>
<b>List of Figures.....</b>	<b>x</b>
<b>List of Tables.....</b>	<b>xi</b>
<b>Chapter One - Introduction .....</b>	<b>1</b>
1.1 The background to the study.....	1
1.2 The statement of the problems.....	3
1.3 Systemic Functional Linguistics .....	4
<i>1.3.1 Language as a social semiotic .....</i>	<i>4</i>
<i>1.3.2 Social functions of language.....</i>	<i>5</i>
1.4 The aims of the study and the research strategy .....	8
1.5 The significance of the study .....	10
1.6 The overview of the thesis.....	11
<b>Chapter Two - Rubrics, training of markers and the reliability of assessment .....</b>	<b>13</b>
2.1 Recent studies in assessment.....	13
<i>2.1.1 The reliability of assessment .....</i>	<i>15</i>
2.2 Rubric and reliability .....	17
<i>2.2.1 Holistic rubric .....</i>	<i>20</i>
<i>2.2.2 Rubric use and rater training .....</i>	<i>22</i>
<i>2.2.3 Analytic rubric.....</i>	<i>25</i>

2.3 Rater and reliability .....	28
2.3.1 Raters' language backgrounds and foci of severity.....	30
2.3.2 Rater types .....	32
2.3.3 Raters' perceptions and practices.....	34
2.3.4 Analysis of perceptions using the Appraisal taxonomy .....	37
2.4 Summary .....	44
<b>Chapter Three - Systemic Functional Linguistics .....</b>	<b>46</b>
3.1 Language in context.....	47
3.1.1 The cultural context of language .....	48
3.2.2 The situational context of language .....	57
3.2.3 Language strata and systems .....	61
3.2.4 The clause complex system.....	63
3.2.5 Theme.....	65
3.2.6 Appraisal theory.....	66
3.3. Summary .....	70
<b>Chapter Four - Methodology.....</b>	<b>71</b>
4.1 Research purpose .....	72
4.2 Research design.....	73
4.2.1 Qualitative approach.....	73
4.2.2 An action-oriented case study.....	74
4.2.3 Participant selection.....	75
4.3 Data collection.....	78
4.3.1 Phase one – Exploring exiting practices.....	80
4.3.2 Phase two – Redesigning practice.....	81
4.3.3 Phase three – Implementing new practices.....	85

4.3.4 Essay marking.....	86
4.3.5 Ethical considerations.....	87
4.3.6 Research relationships.....	88
4.4 Triangulation of data.....	88
4.5 Data analysis tools.....	90
4.5.1 Consensus estimate of inter-rater reliability.....	90
4.5.2 Linguistic analysis.....	92
4.6 Summary .....	108
<b>Chapter Five - Designing SFL-informed rubrics .....</b>	<b>109</b>
5.1 Analysis of two sample essays .....	109
5.1.1 Genre analysis.....	110
5.1.2 Analysis of the ideational function of text: the clause complex system .....	115
5.1.3 Lexical density and nominalisation.....	124
5.1.4 Analysis of the textual function of text: the Theme system .....	129
5.1.5. Analysis of the interpersonal function of text: the Appraisal system .....	138
5.2 Establishing an SFL-informed rubric .....	143
5.3 Training teachers.....	148
5.3.1 The training cycle.....	148
5.3.2 Training sessions and activities.....	151
5.4 Summary of the chapter .....	159
<b>Chapter Six - Teachers' perceptions and practices .....</b>	<b>161</b>
6.1 Teachers' opinions and practices: Pre-intervention .....	161
6.1.1 Teachers' opinions of the existing marking rubric.....	164
6.1.2 Teachers' opinions about their consistencies and practices.....	172
6.1.3 Pre-training inter-rater reliability and sources of inconsistency .....	179

6.2 Teachers' perceptions of the SFL-informed rubric .....	191
6.2.1 Teachers' opinions of the SFL-informed rubric.....	193
6.2.2 Teachers' opinions about their consistencies and practices when using the SFL- informed rubric .....	202
6.2.3 Teachers' opinions of the intervention.....	204
6.2.4 Post-training inter-rater reliability .....	208
6.3 Summary .....	224
<b>Chapter Seven - Discussion and conclusion .....</b>	<b>226</b>
7.1. The relationship between teachers' perceptions and their assessment practices	227
7.1.1 The influences of the linguistically informed assessment criteria on inter-rater reliability.....	228
7.1.2 The importance of understanding teachers' perspectives.....	230
7.1.3 Teachers' opinions and their marking practices.....	234
7.1.4 Teachers' perspectives and marking ethics .....	239
7.1.5 Improved teachers' linguistic knowledge .....	241
7.1.6 Summary.....	246
7.2 Conclusions and implications.....	247
7.2.1 Implications for assessment practice .....	248
7.2.2 The contributions of this thesis .....	250
7.2.3 Limitations of the study.....	252
7.2.4 Recommendations for further research .....	252
7.2.5 Concluding remarks .....	253
<b>References.....</b>	<b>255</b>
<b>Appendix B: Consent Form.....</b>	<b>276</b>
<b>Appendix C: Interview Guide .....</b>	<b>278</b>

<b>Appendix D: colour coding and abbreviation.....</b>	<b>280</b>
<b>Appendix E: Training Materials .....</b>	<b>281</b>
<b>Appendix F: Genre analysis of essay 8a and 10a .....</b>	<b>306</b>

## List of Figures

Figure 2.1 Holistic rubric .....	18
Figure 2.2 Analytic rubric.....	19
Figure 2. 3 Model of Appraisal resources .....	39
Figure 3.1 Context of culture.....	50
Figure 3.2 Map of genres in schooling .....	53
Figure 3.3 Stages and phases of exposition .....	56
Figure 3.4 Stages and phases of discussion .....	57
Figure 3.5 Context of situation: register (field, tenor and mode) .....	58
Figure 3.6 Register variables and metafunctions correlation .....	60
Figure 3.7 Language strata .....	62
Figure 3.8 The context relationship with the language functions and systems .....	63
Figure 4.1 Phases of data collection.....	78
Figure 4.2 Teachers' marking reliability when using the existing rubric.....	92
Figure 4.3 Layers of Themes .....	99
Figure 5.1 macroTheme and hyperTheme development in teacher-rated essay .....	131
Figure 5.2 Teacher-rated essay: Thematic progression in the Thesis stage .....	133
Figure 5.3 Teacher-rated essay: Thematic progression in the Argument 1 stage .....	135
Figure 5.4 Teacher-rated essay: Thematic progression in the Argument 2 stage .....	136
Figure 5.5 Thematic progression in the Restatement stage.....	138
Figure 5.6 Teachers' perceived problems and level of agreement in terms of criteria used in the existing marking rubric .....	147
Figure 5.7 The teaching cycle.....	149
Figure 6.1 Attitudinal choice: positive and negative perceptions .....	163
Figure 6.2 Attitudinal choice: positive and negative perceptions (post-training) .....	193
Figure 6.3 Teachers' consistency before and after the training .....	210
Figure 7.1 A summary of teachers' perceptions of the two rubrics.....	233
Figure 7.2 Diagrammatic illustration of comparative shift in teachers' marking practices.....	245

## List of Tables

Table 3.1 A classification of key arguing genres .....	54
Table 3.2 Four common arguing genres in student academic writing .....	55
Table 4. 1 The links between purposes/research questions and methods .....	72
Table 4.2 The research participants .....	77
Table 4.3 The existing marking rubric.....	79
Table 4. 4 Interview excerpts on difficulties in marking vocabulary .....	83
Table 4. 5 Theme-Rheme pattern .....	96
Table 5.1 Clause combination in the IELTS-rated essay .....	116
Table 5.2 Clause combination in the teacher-rated essay.....	118
Table 5.3 Nominal groups in grammatical metaphor .....	123
Table 5.4 Lexical density and nominalisation in the two essays.....	125
Table 5.5 Distribution of Appraisal resources in the two essays .....	141
Table 5.6 Score traits of the SFL-informed rubric .....	144
Table 5.7 Genre .....	145
Table 5.8 Register .....	145
Table 5.9 Accuracy.....	146
Table 5. 10 Activities conducted during the intervention program.....	148
Table 5.11 The interpersonal and the textual functions of language .....	152
Table 5.12 Social purposes and schooling genres .....	153
Table 5.13 Genre and register concept activity .....	155
Table 5.14 Genre and register concept activity: sample of discussion essay .....	155
Table 5.15 Exercise 5: Appraisal resources finding activities .....	157
Table 6.1 Teachers' positive-negative perceptions of the existing marking rubric.....	164
Table 6.2 Teachers' use of Judgement in terms of capabilities and practices .....	173
Table 6.3 Pre-training consensus estimates of inter-rater reliability .....	180
Table 6.4 Rate of recurrence of the ten essays.....	181
Table 6.5 Score gap in terms of genre stages .....	183
Table 6.6 Score gap: accuracy features of essays 1 and 7.....	185
Table 6.7 Subject-Verb Agreement in Essay 1.....	187
Table 6. 8 Teachers' positive and negative perceptions of the SFL-informed rubric...	194
Table 6. 9 Teachers' use of Judgement in terms of capabilities and practices when using the SFL-informed rubric.....	202
Table 6.10 Teachers' positive and negative perceptions of the intervention.....	204
Table 6.11 Post-training consensus estimate of inter-rater reliability .....	209
Table 6.12 Classification of student essays: genre, mean score and frequency.....	211
Table 6.13 Score gap of Essay 8a and Essay 10a based on genre stages .....	215
Table 6.14 Score gap of Essays 8a and 10a based on register variables.....	218
Table 6.15 Score gap of Essay 8a and Essay 10a based on accuracy of the essay .....	223

# Chapter One

## Introduction

### 1.1 The background to the study

In our global society, the ability to write effectively is becoming increasingly more important. The increasing role of writing in both second- and foreign-language education has raised an ever-greater demand for valid and reliable ways to test (to assess) writing ability, both for classroom use and as a predictor of future professional or academic success (Weigle 2002). This demand for valid and reliable ways of assessing writing ability has been a problem for me, as the researcher, who has been engaged in the teaching of English for Academic Purposes (EAP) program at a language institution in Indonesia for approximately eight years.

In the language institution where I worked, one EAP class is normally taught by two or three teachers, depending on whether the class is held twice or three times a week. To help us with marking our student essays, the school provides us with assessment criteria (a marking rubric) for each student assignment. The rubric consists of eight criteria which mainly focus on the marking of students' essays at the micro level (e.g., *Subject – Verb Agreement*, *Spelling* and *Punctuation*) and the clustering of different aspects of writing skills into one criterion (e.g., *Vocabulary* is combined with *Spelling* and *Punctuation*). The problem appears that despite the marking criteria provided by the institution, there existed inconsistencies in our marking of student essays. One teacher can be too lenient and tend to assign high scores while the others tend to be too strict in their marking. The discrepancy of

scores between us, the teachers, not only happens because of the different degree of strictness or leniency of the teachers but also due to our misinterpretations of the criteria used. Based on the marking experience and the initial conversations I had with my colleagues, it appeared that there were some aspects of writings that were not covered in the existing rubric we used, and it was quite often that we perceived one criterion differently. This has urged the need to conduct an intervention that can help develop more systematic assessment criteria that become shared knowledge among the teachers. It was expected that training the teachers to use the new rubric would promote a higher level of consistency in teachers' essay-marking practices. This was the motivation that guided and controlled the direction of the study.

The present study examines the effect of SFL-informed assessment rubric and rater training on the reliability of teachers' marking of student writing. The literature has demonstrated the importance of rater reliability in the testing or assessing of writing skills (East 2009, Weigle 2002). Inferences of students' language ability are always made based on the results of testing or assessing, and decisions such as admission to academic programs or placement into different levels of a language program, or strengths or weaknesses of individual students, are always made based on those inferences (Weigle 2002). Despite the difficulty in achieving rater reliability in tests of writing skills due to the complex nature of writing proficiency (H. D. Brown 2004, Knoch 2009), some research has shown that some careful specification of an analytical scoring and rater training, particularly ongoing training, can help in improving rater reliability (e.g., Barkaoui 2007, Hamp-Lyons 2007, Weigle 1994; 1998). As Systemic Functional Linguistics (SFL) offers a range of tools for

systematic analysis of texts, the present study argues that a functionally oriented assessment rubric will help improve rater reliability.

## **1.2 The statement of the problems**

Two main problems have been identified in relation to this study. The first is the inconsistency in teachers' rating of student essays. The second one is the unavailability of clear and effective marking criteria for the students' assessment tasks. Thus, standardisation of assessments of student writing is at stake.

As writing is highly valued in educational settings, the standardisation of writing means that correctness or convention in writing is frequently regarded as being more important than correctness in speaking (Weigle, 2002). The importance of convention in writing is particularly relevant for writing in academic contexts, in which writing is frequently seen as a key to entry into the 'academic discourse community' (Spack 1988, Swales 1990, Weigle 2002). This, however, has not been reflected by the accuracy of raters in their assessment of students' academic writing despite the marking criteria writing provided. In other words, it can also be said that the raters' attempts to rate student essays more objectively and accurately have not thus far been reflected in rating scale descriptors (Knoch 2007; 2008). Thus, the second problem, which is the unavailability of simple, clear yet effective scale descriptors, is grounded upon this claim. As Brown (2004) argues, the difficulty in achieving rater reliability in tests of writing skills is mainly due to the complexity of traits represented in writing proficiency, which are difficult to define. However, the careful specification of an analytical scoring instrument informed by linguistics,

such as the SFL informed assessment, can increase rater reliability (J. D. Brown 1991).

## **1.3 Systemic Functional Linguistics**

Systemic Functional Linguistics (SFL) is a linguistic theory developed by Michael Halliday and his colleagues (e.g., Halliday 1985a, Halliday 1994, Halliday and Matthiessen 2004, Martin 1992). It is concerned with the notion of language as a social semiotic or resource for making meaning in socio-cultural activities such as education.

### **1.3.1 Language as a social semiotic**

The first part of the notion of social semiotic is “social”. The term “social” in the SFL concept of language suggests two things simultaneously. One is “social” used in the sense of the social system, which is synonymous with the culture. The other is the term “social” used to indicate that SFL is concerned particularly with the relationships between language and social structure, considering the social structure as one aspect of the social system (Halliday and Hasan 1985). Eggins (2004), uses the term ‘functional-semantic approach’ to indicate that SFL theory explores how people use language in different contexts. Because of the importance of the context of culture, which includes institutions such as language schools and context of situations such as assessment tasks, this theory is useful for exploring language use in a range of settings, including an EFL academic writing course such as in Indonesia.

The second part of the notion is “semiotics”. Semiotics in its most general sense can be defined as the study of making and understanding meaning (Halliday and Hasan 1985, Halliday and Matthiessen 2004, Lemke 1995). SFL is concerned with exploring how language is structured for use as a system of making and understanding meaning.

### 1.3.2 Social functions of language

According to the SFL theory, language serves three kinds of social functions (known as “metafunctions”): to talk about one’s experience of the world, to interact with other people and to organise message. These are referred to as the **ideational**, **interpersonal** and **textual** metafunctions, respectively (Halliday 1994). Drawn from his study on his son’s language development, Halliday’s major contribution has been his argument on systematic correlation between the three metafunction and specific contextual features. Halliday (1975; 1978; 1985a) goes as far as saying that each function interacts with variables of context (**field**, **tenor** and **mode**, where field focuses on “the subject matter”, tenor on the “role relationship between interactants”, and mode on “the means of interaction”).

The ideational metafunction refers to how language users express their experiences, perceptions or ideas about the world and how they connect these ideas together. It represents the speaker's meaning potential as “an observer”. It is the content function of language, through which the speaker or writer use language to encode their experiences as a member of culture to make sense of the world (Halliday 1978, Martin and Rothery 1993). This metafunction consists of the systems of transitivity, clause complex and nominalisation. In educational settings, such as in a language

school where the study was conducted, an analysis of grammatical resources (i.e., process, participant and circumstance types) and clause complexing (how clauses are formed and combined) used in the text books, writing samples, student writings and teachers' talk would reveal how the 'world' is viewed by the speakers or writers.

The interpersonal function refers to the way in which language users draw on language resources to construct and maintain relationship with others. In other words, this is the component through which they intrude themselves into the context of situation, by both expressing their own attitudes and judgements and seeking to influence the attitudes and behaviours of others. It expresses the role relationships associated with the situation, including those that are defined by language itself, relationships of questioner-respondent, informer-doubter and the like (Halliday 1978, Martin and Rothery 1993). Part of the interpersonal function relevant to this study is the system of Appraisal, which was developed by Professor James Martin of University of Sydney (White 2001; 2012). It has emerged from within SFL and is a semantic tool in the SFL framework used for the analysis of evaluation by speakers/writers, and helps with exploring, describing and describing the way language is used to evaluate, to adopt stances and to manage interpersonal positioning and relationships (White 2002; 2012). In an academic writing course, for example, the interpersonal function can be identified in student essays through choice of words used to express feelings, evaluate an entity or make judgements of a person, as well as through the way they try to influence or convince the readers.

The textual metafunction is a set of resources that allows us to produce and organise texts that make sense in a context of situation. It refers to meaning about how language users organise their intended messages so that these messages are “cohesive, coherent and well-crafted” (Christie and Derewianka 2008, p. 24). It is the enabling function of language that represents the speaker’s text-forming potential. This metafunction expresses the relation of the language to its environment, including both the verbal environment – what has been said or written before – and the nonverbal, situational environment. The function is concerned with the organisation of a communication, constructing symbolic reality as a wave of information (Halliday 1978, Martin and Rothery 1993). This function is realised through the system of Theme. This system describes the structural configurations by which the clause is organised as a message. In English texts, clauses typically begin with something that is familiar or already known to the reader and then moves on to something new. The part of the clause that is the point of departure of a message is called *Theme* (Halliday 1994, Halliday and Matthiessen 2004) and the rest of the clause in which the Theme is developed is called *Rheme* (Eggins 1994 , p.275).

The three metafunctions (the ideational, interpersonal and textual) are realised through systems of choices or options at the level of clause and at the level of text. They are “the modes of meaning that are present in every use of language in every social context” (Halliday 1978, p. 112) drawn on by language users as they interact via spoken and written language in a language school (college) setting, including text books, seminars, lectures and of particular interest to the present study, student essays. For example, student essays will reveal students’ understanding of certain topics such as globalisation, immunisation and its impacts on developing countries

(the ideational metafunction), the attitudes and judgments they make regarding the topic, how they use words to amplify these attitudes and how they try to influence their readers (the interpersonal metafunction) as well as how the message or information relating to globalisation and its impacts is organised to achieve the purpose (the textual metafunction).

## **1.4 The aims of the study and the research strategy**

The aim of the study was twofold. Firstly, it aimed to enhance the consistency of the teachers' essay marking by developing the linguistic-based marking criteria, informed by Systemic Functional Linguistics (SFL) developed by Halliday and colleagues (Halliday 1978; 1994, Halliday and Hasan 1985, Halliday and Matthiessen 2004, Martin 1984). Secondly, it aimed to examine the essay marking experiences of English for Academic Purposes (EAP) teachers at a private English course in Indonesia when using the marking rubric provided by the institution and when using the SFL-informed marking criteria.

An action-oriented case study was selected for the investigation of teachers' marking practices. As the name suggests, this action-oriented research aimed to achieve a profound understanding of people in the real world and to improve practice or develop individuals (Herr and Anderson 2005, Pihlanto 1994). In this study, the case was a group of EAP teachers' marking performance and practices. This research strategy was chosen because it allowed me to investigate a case such as teachers' descriptions of their assessment practices and consistencies and to make

an action by designing a linguistically informed marking rubric that was introduced through an intervention. Thus, this study addressed the research questions below:

1. How do raters perceive the existing marking criteria?
  - What do raters identify as the most difficult or problematic aspects of student texts to mark?
2. What are the main sources leading to rater inconsistency in student essays?
  - What features of student texts appear to be least agreed upon?
  - What features of student texts are there most agreement about?
3. How do raters' marking practices shift after using SFL-informed criteria?
4. How do raters respond to the SFL-informed criteria?

In response to the first two questions, the study investigated the teachers' perceived problems about their assessment practices in relation to the marking rubric used using a system of Appraisal (Martin and Rose 2007, Martin and White 2005) as a tool of analysis. After that analysis of the teachers' actual marking practices was conducted to identify features of student essays that teachers tended to agree and disagree about, new marking criteria informed by SFL theory were designed by incorporating teachers' comments about their marking difficulties and lack of the existing rubric. These new marking criteria were introduced to the teachers through an intervention (a four-week teacher-training program). The third and fourth questions examined the outcome of the intervention in terms of teachers' evaluative comments about the SFL-informed rubric, the intervention and their perceived consistencies using the system of Appraisal. The teachers' actual marking consistencies, both before and after the training, were measured by using a consensus estimate of inter-rater reliability counts proposed by Stemler (2004).

## **1.5 The significance of the study**

This study is significant for its practical, theoretical and methodological contributions. Practically this study is significant for English as a foreign language (EFL) teachers because it investigates the impact of an SFL informed rubric and training program on inter-rater reliability. The utilisation of SFL in the marking rubrics and text analysis is expected to contribute to assessment practice, especially in rater training because of its potential to increase rater reliability in marking student essays and also its explicitness about successful texts in academic writing. Another practical contribution of this study is that it develops a comprehensive, theoretically consistent and linguistically explicit marking rubric within an SFL framework. As a social semiotic tool, SFL can analyse texts on extra-linguistic levels (contexts of situation and culture), beyond sentence level, as the traditional grammar does. In the Indonesian context, for example, the marking of students' essays are mainly focused on the linguistics levels such as grammar, vocabulary, spelling and punctuation, and teachers tend to have lack of knowledge of why one essay is better than the other in terms of its rhetoric and content. The last practical contribution is that this study focuses to train English teachers by applying an SFL model in an EFL context (Indonesia) where not too many teachers are familiar with SFL as opposed to other countries such as China. Thus, the Indonesian teachers' shared knowledge of the features of Systemic Functional Linguistics, which is incorporated in the new marking rubric, will provide them with tangible textual evidence of judgements of student essays. This can ensure that students' work is marked more fairly and consistently. This shared knowledge also helps them focus on enhancing students' writing skills. In addition, the explicitness of the assessment

criteria can inform learning; it will enable students to see their achievement or progress, and show at which parts they should work more. Theoretically, the utilisation of SFL in assessment marking rubric is expected to enrich the application of SFL theory, which is predominantly used in English-speaking countries in English as a Second Language (ESL) contexts. Its application to assessment of academic writing in an Indonesian educational context would contribute to a wider application of the systemic linguistic theory in assessment in an EFL context, in which such an approach has rarely hitherto been attempted. Methodologically this study uses action research to compare teachers' perception and identify the effects of using an SFL informed model. Many studies in an ESL/EFL context have been conducted on the cross-cultural differences in academic literacy using an SFL model (Appraisal) but little research has attempted to prove it using action research. The study establishes a robust assessment tool to capture vocabulary and expressions by modelling the marking criteria using Appraisal theory to capture critical components which are considered to be the most important criteria. It also analyses teachers' perceptions by adopting Appraisal analysis (not thematic or content analysis) and match the results with the one by non-linguistic analysis of raters' consistency.

## **1.6 The overview of the thesis**

This chapter has provided the overview of the study. The remainder of the thesis consists of a further six chapters. In Chapter Two, relevant literature is reviewed. It focuses on current research on the assessment issues in academic writing such as rater reliability, rater bias, rubric and reliability, and raters' perceptions and practices. Chapter Three gives a detailed account of the theoretical framework which

informed the study. It starts off with an explanation of SFL theory and its key tenets of genre and register, followed by analytical tools to be used to interpret the data. Chapter Four provides a detailed account of the methodology and inquiry processes used in the study. Chapter Five presents the analysis of what was valued in successful student texts to be utilised to inform the design of the SFL-informed criteria and intervention. Chapter Six presents and discusses the findings in relation to teachers' perceptions and marking practices. The first section of the chapter presents teachers' opinions and their actual marking consistencies when using the existing marking criteria provided by the school they work in. The second section focuses on teachers' opinions and their actual assessment practices after the intervention, when using the SFL-informed rubric. The final chapter, Chapter Seven, discusses the key findings in light of the given theory and the relevant literature before drawing some conclusions about the implications of the study. It also suggests possibilities for future research.

## **Chapter Two**

### **Rubrics, training of markers and the reliability of assessment**

The previous chapter has introduced the study by describing the general background and the problem motivating the study, along with the aim and the research strategy. To reiterate, the key concern of the study is to enhance assessment practice surrounding the academic essay by developing more explicit criteria. Such criteria, as it is argued, need to be based on actual linguistic features, which are part of the shared knowledge of teachers, and the importance of ensuring that teachers have a shared understanding. This chapter examines relevant literature in two sections. The first section deals with recent studies of assessment that are concerned with reliability and consistency in marking. The second section deals with studies about the training of markers.

#### **2.1 Recent studies in assessment**

In the real case of an actual barometer we can determine whether the degrees are rightly marked by reference to a constant standard – the yard measure which is kept in the tower. In determining the correctness of the scale according to which an examiner marks, we have not this resource... How far is it likely that any two examiners will differ from each other in their numerical estimate of the same work? What is the probability that there will occur a difference of any assigned extent between two marks given by a pair of examiners to the same work? (Edgeworth 1890, pp. 461-462)

The reliability of raters has been of interests to scholars since the late 19<sup>th</sup> century. It is crucial because test scores have significant impacts on learners' opportunities, institutional resources and public policy (Brossell 1986, East 2009). As a result, much effort has gone into the development of efficient and reliable measures of

assessment. From the seminal work of Edgeworth (1890), who as we have seen above, lamented the unavailability of a means to measure the consistency of marking at the time, assessment is claimed to have been plagued by concerns about the reliability of rating, usually among markers (Hamp-Lyons 2007).

Assessment, in the educational context, can be defined as ‘all sorts of activities that teachers engage in to evaluate their students’ progress, learning needs, and achievements’ (Weigle 2007, p. 195). Assessment practices in the educational context cover different fields such as science, sport and language, and are evident in both low and high-stakes tests. In Australia, for instance, the high-stakes assessment practice like the National Assessment Program – Literacy and Numeracy (NAPLAN) and the Higher School Certificate (HSC) are considered obligatory for all Australian students at certain levels (i.e., Years 3, 5, 7 and 9 for NAPLAN and Years 11 and 12 for HSC). Prior to assessing students’ work, markers for these assessment practices are trained to use the criteria-based measures to ensure the consistency and reliability of the scores assigned.

In general, high quality assessments are considered to be those with a high level of reliability and validity. Reliability is concerned with consistency and dependability of the assessment (Feldt and Brennan 1989, Gay 1985), whether the results of assessment would be similar if conducted by the same person in different time or occasions. Validity is the extent to which inferences made from assessment results are appropriate, meaningful and useful in terms of the purpose of assessment. In other words, a valid measurement assesses what it claims to assess (Bachman 1990, Charney 1984, Hughes 2003). Establishing the reliability of assessment is important

because it is seen as a prerequisite for establishing validity (Gay 1985, P. Winke 2011). Although a valid assessment is by necessity reliable, the contrary is not true. A reliable assessment is not necessarily valid. If an assessment is measuring what it is supposed to measure, it will be reliable and do so every time. However, a scoring rubric is likely to result in invalid interpretations, when the scoring criteria are focused on an element of the response that is not related to the purpose of assessment (Moskal and Leydens 2000).

### **2.1.1 The reliability of assessment**

Maintaining the reliability of assessment, particularly when the tests require answers or responses that are not inherently countable, is somewhat challenging. Murphy (1982) conducted a four-year investigation into the reliability of marking of a variety of Associated Examining Board (AEB) examinations at Ordinary level (O-level) and Advanced level (A-level), General Certificate of Education (GCE) examinations. The examination papers consisted of different types of responses to questions such as: short-answer, structured, free-response, and essay questions. The level of reliability in this four-year study was computed on a mark and re-mark basis, i.e., after the completion of the normal marking period, the examination scripts were prepared for remarking by having all previous marks and examiners' comments removed from them. The results were analysed in terms of correlations between original marks and the re-marks. The study demonstrates that the least reliably marked examinations tend to be those that place the most dependence on essay-type questions. This affirms Cast's (1939) view of the assessment of essays as one of the most unreliable modes of mental measurement (p. 257). In academic writing courses such as EAP programs, for example, the correctness of the score

assigned to an essay cannot be simply determined by straightforwardly measured criteria, such as grammatical and mechanical aspects of the essay (i.e., grammatical accuracy, punctuations, spelling and word counts), as there are other less straightforwardly measured criteria that should be taken into consideration (such as coherence of ideas, content and vocabulary). In terms of vocabulary, for example, although it is of particular importance for second and foreign language writers and is often assessed in ESL/EFL writing, raters reliability to assess the range and accuracy of vocabulary usage is yet undetermined (Ruegg, Fritz and Holland 2011). This often results in the inaccuracy of scores assigned.

The reliability of assessment is determined by the consistency of scores assigned. The more consistent the scores are over different assessors and occasions, the more reliable the assessment is thought to be (Moskal and Leydens 2000). In this sense reliability is synonymous with replicability (Charney 1984, Gipps 2012). In the assessment of a student essay, for example, the score is considered reliable if it is replicable. That is, the same score occurs when marked by the same person on a different occasion. When we speak of reliability, we are also referring to the score consistency of a group of assessors (Gay 1985), i.e., whether the essay will receive similar scores if marked by two or more raters. To maintain reliability of assessment, the primary question that should be answered is the factors that influence the consistency or inconsistency of scores assigned by the assessors. It is important to examine the influence of marking rubrics, language background, and rater training on the reliability of markers. Similarly, it is of the same importance to examine the thoughts of the assessors involved in the marking i.e., their thoughts about the marking rubric used and their marking capabilities. These are the

questions driving the direction and locating the position of the present study in the literature.

Literature on rater reliability reveals two major themes. The first is concerned with the types of rubrics used to assess the students' work. Some of the examples are: East's (2009) research on the reliability of analytic rubric, Barkaoui's (2007) researching into the positive impacts of holistic rubric and Boettger's (2010) exploration of holistic and analytic rubrics use in technical communication. The second theme is concerned with assessors as influencing factors in the assessment practices. This includes but is not limited to the work of Eckes (2008), Kondo-Brown (2002) and Schaefer (2008), who investigated rater bias patterns in the assessment of writing, Hamp-Lyons and Zhang's (2001) and Johnson and Lim's (2009) research which examined the influence of rater language background on assessment of writing. To continue the second theme, some other research is concerned with the assessors' language background with their severity and leniency in assessing academic essays (e.g., H. K. Lee 2009a, Shi 2001) and speech performance (e.g., Gui 2012, P. Winke, Gass and Myford 2013). It is from these themes that the literature to be reviewed is selected and discussed.

## **2.2 Rubric and reliability**

The influence of scoring rubrics on the reliability of rating has been of considerable interests to researchers. Rubric, which originally means "mark in red", is a term probably derived from the past traditions of highlighting or marking papers with red ink (Finson and Ormsbee 1998). In the educational literature and among teaching

and learning practitioners, the word ‘rubric’ can be defined as “a simple assessment tool that describes level of performance on particular task and is used to assess outcomes in a variety of performance-based context” (Hafner and Hafner 2003, p. 1509). In America, for example, rubric was first defined for writing assessment in the mid-1970s (Dirlam and Byrne 1978) and used to train raters for New York State Regent Exam in Writing by the late 1970s for the purpose of establishing standardised ratings (Dirlam 1980). As an attempt to communicate expectations of a quality around a task, a rubric, in many cases, is used to delineate consistent criteria for grading. However, the delineation of consistent criteria does not inevitably result in consistency of the scores assigned due to the subjectivity in the rater’s evaluation (Elder, Barkhuizen, Knoch and von Randow 2007). In other words, along with the consistent marking criteria, consistency in marking can be made possible if the raters’ appropriate and consistent use of the marking scales can be ensured.

Thereafter, major issues concern the types of rubric scales used in assessing written work. Two of the most commonly used are holistic rubric (see Figure 2.1), which assigns a single score to a script based on the general impression of the script and analytic rubric (see Figure 2.2), which rates scripts based on several aspects of writing or criteria (Hamp-Lyons 1991a, Weigle 2002).

SCORE	RESPONSE DESCRIPTION
5	The need, solution, example, and the required amount of the specific amount are identified.
4	The need, solution and either the example or required amount of the example are identified correctly.
3	The need and the solution are identified correctly. The example is inaccurate <u>and</u> the amount of the example is incorrect.
2	Both the solution and the required amount of the example are identified. The example is inaccurate <u>and</u> the amount of the example is incorrect.
1	Only the example <u>or</u> the required amount of the example is identified correctly. The need and the solution are not identified or they are incorrect.

**Figure 2.1 Holistic rubric (Silvestri and Oescher 2006, p. 30)**

ESL COMPOSITION PROFILE				
STUDENT	DATE		TOPIC	
SCORE	LEVEL	CRITERIA	COMMENTS	
CONTENT	30-27	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic		
	26-22	GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail		
	21-17	FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic		
	16-13	VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate		
ORGANIZATION	20-18	EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/ supported • succinct • well-organized • logical sequencing • cohesive		
	17-14	GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing		
	13-10	FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development		
	9-7	VERY POOR: does not communicate • no organization • OR not enough to evaluate		
VOCABULARY	20-18	EXCELLENT TO VERY GOOD: sophisticated range • effective word/ idiom choice and usage • word form mastery • appropriate register		
	17-14	GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage <i>but meaning not obscured</i>		
	13-10	FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • <i>meaning confused or obscured</i>		
	9-7	VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate		
LANGUAGE USE	25-22	EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions		
	21-18	GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions <i>but meaning seldom obscured</i>		
	17-11	FAIR TO POOR: major problems in simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • <i>meaning confused or obscured</i>		
	10-5	VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate		
MECHANICS	5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing		
	4	GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing <i>but meaning not obscured</i>		
	3	FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • <i>meaning confused or obscured</i>		
	2	VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate		
TOTAL SCORE	READER		COMMENTS	

Figure 2.2 Analytic rubric (Jacobs, Zingraf, Wormuth, Hartfield and Hughey 1981, cited in Weigle, 2002, p. 116)

For example, Bauer (1981) and Barkaoui (2007) compared the reliability of analytic and holistic scales in assessing writing. Tedick and Mathison (1995) investigated the correlation between holistic scoring and the rhetorical features in ESL writing. East (2009) evaluated the reliability of analytic scoring rubric for foreign language writing. Although it is argued that “perfectly reliable criterion measure exists only in theory” (Coffman 1966, p. 153), the positive effects of using rubrics to guide and promote the reliability of scores is unquestionable (see, for example the work of Barkaoui 2007, East 2009, Hafner and Hafner 2003 among others, which will be reviewed afterwards). The discussion below outlines the literature on the importance of rubrics in relation to rater-reliability.

### **2.2.1 Holistic rubric**

Assessment practices tend to result in higher reliability when the rubric is used (Jacobs, Zingraf et al. 1981, Silvestri and Oescher 2006). One of the studies comparing the reliability of assessors when marking with and without rubrics was done by Silvestri and Oescher (2006). In the study, 16 pre-service teachers were asked to assess four sample papers on students' knowledge of health-related issues without using a rubric, and a second time with a holistic rubric developed by the researchers (see Figure 2.1). On the first occasion, the teachers were requested to rate each paper on a scale of 1 to 5 with the latter score representing the highest possible score. After a short break, the participants were provided with the rubric and the criteria were explained. Then, they were asked to assess the same papers using the rubric given. In addition, each of the teachers was asked to reflect on two questions related to (1) opinions about the differences between marking with and without rubrics and (2) preference between the two marking methods and the reasons for their preferences. The study revealed that the teachers' reliability was significantly higher when marking with a rubric, and the majority of them indicated the use of rubric resulted in consistent grading that was far less subjective in nature.

Holistic scoring is widely used in writing assessment due to its practicality, and with appropriate training, it can result in high levels of score reliability (Vacc 1989, Weigle 2002). Barkaoui's (2007) comparative analysis on the effect of rating scales on raters' performance reveals that holistic scales tend to result in higher inter-rater agreement and practicality. It is claimed as practical because the marking can be done faster as the raters can rely on the impressionistic criteria when choosing one score to reflect the quality of the essay being marked. However, his study concludes

that assessors are the main source of variability in marking. As Barkaoui (2007) admitted as the limitation of his study, it is uncertain whether the findings with regard to assessors would hold if rater training were provided as suggested by other previous studies (such as, Grabe and Kaplan 1996, Rezaei and Lovorn 2010, Vaughan 1991).

Despite claims about the reliability of holistic scales, some researchers argue that variation is a common feature in the marking of sub-skills or criteria. In essay-type assessments, these sub-skills include aspects commonly labelled Grammar, Vocabulary, Content, Organisation, Punctuation and others. Brown (1991), in a study of inter-assessor score agreement, investigated the degree to which differences existed in the writing scores given by English and ESL faculties. Eight teachers, each from English and ESL faculties, rated 112 student essays using a holistic 6-point (0-5) rating scale. The study found that the scores given by the two groups of teachers were statistically similar. Despite the similarity of scores assigned by the two groups, Brown (1991) argues that the teachers may have arrived at those scores from somewhat different perspectives. In other words, holistic scoring is unfavourable because assessors do not necessarily use the same criteria to arrive at the same scores (Weigle 2002). Similarly, Shi (2001) argues that high inter-rater reliability does not necessarily mean that the teachers agree on what constitutes a good essay. The study examined differences between native and non-native English as a foreign language (EFL) teachers' marking of the English writing of Chinese university students. A total of 46 teachers (23 Chinese and 23 English-background) assessed 10 expository essays using a holistic 10-point scale. The study found that both groups of teachers, despite the relatively high reliability in their ratings of the

essays, differed in their justifications for such scores. The fact that the teachers differed in their qualitative comments, as the study argues, reflects the teachers' differing understandings of what constitutes a good essay. Accordingly, as Huot (1990) concludes, little is known about the way raters arrive at certain decisions in their marking practices. Such findings provide further evidence of the need to investigate markers' beliefs and experiences, as well as their practices.

While issues of rater reliability have dominated the literature on assessment, particularly in academic writing (Huot 1996), there is a dearth of research concerning the perceptions of the raters with respect to the marking criteria and consistency. The present study argues that investigating the views of raters and their consistencies in actual marking practices will enhance understanding of the correlation between the two (that is, perception and practice). Penny, Johnson and Gordon's (2000) study involving 120 essays written by 5<sup>th</sup>-Grade students revealed the high agreement in scores assigned by the two markers using a holistic rubric. However, Penny, Johnson and Gordon (2000) suggested a follow-up study focusing on rater debriefing needed to be conducted so as to produce a clear understanding of how the raters come to decisions in their marking practices. This is what the current study endeavours to do. Listening to and accommodating the raters' opinions of their marking practices can improve the reliability and validity of rating (Crocker 2002, Norris 2008, Ryan 2002).

### **2.2.2 Rubric use and rater training**

There seems to be consensus in the literature on the positive effects of rater training. Training of assessors, in conjunction with the use of scoring rubric, can help to

contribute to the reliability of scoring procedures (Cronbach 1990, Jacobs, Zingraf et al. 1981). As Lumley (2005) argues, it appears to be impossible to separate rating scale from training process that accompanies it. Weigle's (1994) empirical investigation on the effects of training on inexperienced markers of ESL composition, for example, revealed that training could help raters to understand and apply the writing criteria, and it also modified raters' expectations in terms of characteristics of the writers and the demands of the writing tasks. In the study, 16 assessors (experienced and novice) attended a two-hour training and were then asked to mark four compositions using the English as a Second Language Placement Examination (ESLPE) analytic rating scale, which consisted of three subscales (content, rhetorical control and language). The study used pre/post-training design, to analyse raters' consistency in marking the same essays. The data revealed a higher agreement among raters after the training. The verbal protocols showed that training clarified the intended scoring criteria for raters. Another study by Weigle (1998), which compared severity and consistency among experienced and inexperienced markers before and after training, revealed the tendency of the inexperienced markers to be more severe and less consistent than the experienced ones before training. However, although significant differences in severity were still found among raters after training, the consistency had improved for most raters. Her research findings suggest that rater training is more successful in enhancing raters' self-consistency (i.e., intra-rater reliability) than in getting them to give identical scores (i.e., inter-rater reliability). Nevertheless, since the study compared experienced to inexperienced raters, its findings regarding inter-rater reliability cannot be taken as a whole.

Rater training also clarifies and encourages raters to judge according to the intended rating criteria rather than their own (Charney 1984). A study by Stuhlmann, et al. (1999) examined 40 kindergarten and first grade teachers' abilities to use a rubric to rate 20 First Grade writing samples. A group of 23 of the teachers were trained for one hour on how to interpret the scoring dimensions of the rubric, while 17 others were not. The purpose of the study was to investigate whether training of raters to interpret the scoring dimension would improve reliability. The rubric contained six categories (such as *pre-writing/picture*, *story*, *sentence*, *mechanics/spelling*, *punctuation and capitalisation*, and *format*) that reflect major components of a student's writing development. The study concluded that training had a direct impact on teachers' abilities to consistently apply the scoring criteria used in the rubric.

However, even after the raters have been trained on how to apply holistic rating scales, the inconsistencies among raters can still exist due to the idiosyncratic feature of individual raters and their uniquely complex impressions (Lumley 2002), and differences in their rating experience. Barkaoui (2010a) compared the effects of assessors' varied levels of experience with the scores they assigned when marking using holistic and analytic rubrics. Groups of 31 experienced and 29 novice assessors marked a batch of 24 essays. They received a 30-minute training session on using the holistic and analytic rating scales. Then, the essays were marked individually at the assessor's home and there was a span of two weeks between holistic and analytic rating. In addition to marking the essays, the assessors were instructed to provide a brief written explanation for each holistic score they assigned. The study indicates that the experienced and novice raters differed

significantly in terms of the importance they gave to argumentation and linguistic accuracy when rating essays holistically. Therefore, in addition to investigating raters' perceptions of the rubric and training provided, the present study examines the effect of training on the assessors with similar marking experience.

### **2.2.3 Analytic rubric**

A number of studies into the use of analytic rating scales have indicated their efficacy in enhancing rater consistency. The proponents of the rubrics argue that the scoring schemes are preferred over holistic scoring schemes because they provide more detailed information about a test taker's performance in different aspects of writing (Weigle 2002), and allow for more accurate diagnostic reporting of student progress, particularly in the case of differential skills development as reflected in a student's profile (Weir and Shaw 2007). Furthermore, analytic rubrics have the effect of focusing the raters' judgements and thereby leading to greater reliability (Bauer 1981, Davies, Brown, Elder, Hill, Lumley and McNamara 1999, Weir and Shaw 2007). Two of the earliest studies exploring these viewpoints were conducted by Hartog et al. (1936, cited in Weir 1990) and Cast (1939) who examined the relative effectiveness of analytical and general impression marking for assessing English composition among examiners. When marking analytically, the examiners were asked to allot marks based on the mechanical (handwriting, spelling, punctuation, and syntax), literary (range, appropriateness of information and vocabulary) and the logical aspects of composition (the general organisation of ideas as revealed by the unity, complexity, relevance and the sequence of sentences and paragraphs). Their research found that variation between markers was reduced by the analytic method. Similar findings are reflected by studies that followed. They

include Bauer (1981), Hafner and Hafner (2003), Jonsson and Svingby (2007), and East (2009), among others.

In addition to resulting in higher reliability, the analytic rubrics are also effective assessment tools for assessment among students. Baeur (1981) compared the relative reliability and cost effectiveness of the analytic to the holistic and the primary trait scoring methods. Nine markers who were trained to use the three scoring methods scored a number of secondary school students' essays using each scoring method. Data produced in the study were analysed for the inter-rater and intra-rater reliabilities as well as the cost effectiveness in terms of training and grading times for each of the three scoring methods. The study found the analytic scoring method as the most reliable and the holistic scoring method as the most cost effective in assessing a large number of student essays. Similarly, Hafner and Hafner's (2003) study claims rubrics effectively improved reliability between teachers and its usefulness as an assessment tool for peer, group and self-assessment by students. This study is regarded as one of the first studies providing statistical documentation of the validity and reliability of rubric for student peer-group assessment. By the same token, Jonsson and Svingby's (2007) review of empirical research concludes that reliability can be enhanced provided that the rubric is analytic, topic-specific, and complemented by exemplars and/or marker training. Other research such as Stuhlmann, Daniel, Delinger, Kenton and Powers (1999) and Weigle (1994; 1999), have underlined the importance of training to improve rater performance.

One of the latest studies advocating the use of analytical rubrics was conducted by East (2009). The study shows that raters can succeed in yielding consistent scores in

their marking practice by using the analytic rubric. In this study, two experienced markers were asked to assess 47 essays of German as a foreign language using the rubric criteria, which consist of linguistic features grouped into five categories (see Table 2.1). The study revealed a respectably high level of inter-rater reliability. On his own admission, however, the result of his study must be treated with some caution for two main reasons. Firstly, as there were only two raters involved in the marking, the generalizability of the findings is somewhat limited. Furthermore, the evaluation relied primarily on the quantitative data available from the test scores.

As evidence from the studies cited above, rubrics have been proved to have potential to increase agreement between assessors' marking regardless of the types of rubrics used. This suggests that variability of scores assigned does not seem to be influenced by the kinds of rubrics used. As the previous literature has demonstrated, the utilisation of rubrics has led to improved reliability between assessors. If, as the literature suggests, raters are the main source of the variability of scores assigned (Barkaoui 2007) due to different interpretations of criteria (Barkaoui 2010b) and different levels of expertise (Barkaoui 2010a, Cumming 1990, Weigle 1998), then the key solution to improved reliability in teachers' assessment is probably "the provision of clear marking criteria and schemes"(Gipps 2012, p. 137).

However, regardless of the effectiveness of the two kinds of rubrics, as the literature has demonstrated, analytic rubrics seem to provide more detailed information about students' or test-takers' language proficiency on different aspects of writing (Cumming 1989; 1990, Weir and Shaw 2007). The detailed information can add to the depth of the data that can be used to identify specific strengths and deficiencies

within a program (Boettger 2010). Holistic rubrics, although less time-consuming to apply, are not focused on the task assigned (Tedick and Mathison 1995). By focusing on an overall impression of the essay being scored, a holistic rubric does not seem to inform learning, as it does not give detailed information as the analytic rubrics do on the students' capabilities in writing.

To conclude, it is evident from the literature on rubrics and reliability that the use of rubrics could result in improved consistency in marking, and that markers are considered to be the main source of variability of scores. In light of the concerns raised by Swain (1993) and Moss (1994) about traditional interpretation of reliability, it is therefore important to conduct further research promoting an in-depth understanding of raters' decision-making processes when assigning scores. Further qualitative studies on raters' opinions of the issues of reliability such as the rubric used and the training provided would contribute to an in-depth understanding of the issues. The next section reviews relevant studies that have investigated raters in relation to reliability of scores.

## **2.3 Rater and reliability**

In the direct assessment of writing ability, rater effects can contribute to variability in marking (Wiseman 2012). As suggested in the previous studies (see, for example Barkaoui 2007, Lumley 2002, Weigle 1999), raters account for a substantial amount of variance of the scores assigned. Such difference can result from characteristics of raters such as severity and leniency in making judgements, and are regarded as a possible source of measurement errors in performance testing (Bachman, Lynch and

Mason 1995).

One of the possible sources of variability in assessment is rater bias. Kondo-Brown (2002) identified the bias pattern among raters' marking of Japanese L2 writing. The essay writing was an optional part of the Japanese language placement test. In this study markers were found to be significantly biased against candidates whose responses were located at the very extreme; that is extremely high or low. In this study, essay samples from 234 students with a wide range of language proficiency were scored by three raters, who were all experienced Japanese language instructors working at the same university. The assessment employed a modified version of Jacob's analytic rubric consisting of five main skills: content, organisation, vocabulary, language use and mechanics. Before the assessing process the assessors received a total of three hours of training in assessment procedures. One methodological flaw in the study was the short time allocated to train the raters. Even though the assessors involved in study were well experienced, it could be argued that three hours of training was not sufficient to modify what had been internalised in the assessors' marking practices. It is therefore worth investigating whether a more extensive rater training can help reduce the rater bias pattern in assessment practices.

A number of studies into the reliability of raters have also indicated rater severity as a potential source of inconsistencies in assessment practices. Lynch and McNamara (1998) in a study of ESL speaking skills of 83 immigrants argue rater bias pattern as a source of inconsistencies in marking, in which raters were identified to be "more lenient or severe than other raters for certain persons" (p. 166). Similarly, a study by

Schaefer (2008) on the existence of rater bias patterns in EFL writing assessment also corroborates the finding above and concludes that raters tend to be more severe towards higher ability writers and more lenient towards the lower ability ones.

### **2.3.1 Raters' language backgrounds and foci of severity**

Research assessment exploring the relationship between markers' language backgrounds and bias is not undivided. Although some studies such as Brown (1991), Johnson and Lim (2009) and Shohamy et al. (1992) found rater background (i.e., professional EFL-speakers vs. professional native-English speakers) to have little effect on rater reliability, other studies such as Barnwell (1989), Hamp-Lyons and Davies (2008), Hamp-Lyons and Zhang (2001), Hyland and Anan (2006), Lee (2009a), Shi (2001) and Song and Caruso (1996) identified a strong correlation between rater background and severity. In general, non-native speakers have been shown to be more severe in their assessment than the native speakers. For example, in a study of raters' language background and the foci of severity, Hyland and Anan (2006) found that Japanese teachers of English essay writing were generally more severe in grading errors and relied more on rule infringement rather than intelligibility, while the native-English speaking teachers tended to avoid an 'error hunting' approach and saw appropriacy as a basis of error judgements. In the same vein, Lee (2009a) distinguishes raters' different foci of severity. Her investigation on rating behaviour between Korean and native-English speaking raters indicated that Korean raters were more severe in scoring grammar, sentence structure, and organization, while the native-English raters were stricter toward content and overall scores. This finding affirmed the results of an earlier study by Song and Caruso (1996). In the study, English and ESL faculty raters varied significantly in assigning

the holistic scores. While the English faculty gave greater weight to the overall content and quality than to language usage, the ESL faculty tended to focus on grammatical accuracy.

In addition to different foci of severity, shared language background is believed to have influenced raters' objectiveness. An important study by Hamp-Lyons and Davies (2008) showed the presence of raters' favouritism with students who shared language backgrounds with them. Their empirical study on essays written for the Michigan English Language Assessment Battery (MELAB), an international examination of English language proficiency, revealed the impossibility of raters to be totally free from the influences of their language backgrounds. The sample for this study consisted of 60 essays written by native speakers of Bahasa Indonesia/Malay, Chinese, Arabic, Japanese, Tamil and Yoruba. These essays were firstly rated by native English speakers using the official MELAB ratings. Then, they were rated by raters who shared the same L1 with the examinees, and also by raters who did not share their L1. This study concluded that none of the raters was 'totally' free of the influence of their L1s. This confirms the results of an earlier study by Hamp-Lyons and Zhang (2001), which argued for the persistent influence of raters' cultural background. They claimed the tendency of students' writings to be "favoured by raters whose rhetorical patterns share similar cultural and ideological background" (p. 112).

This finding is, however, not fully confirmed by Johnson and Lim's (2009) study. Their investigation into language background-related bias in the assessment of writing performance found language background to have little effect on students'

scores, and therefore concluded non-native English-speaking (NNS) raters may be trained to be just as effective assessing writing performance as are their native English-speaking (NES) counterparts. In the study, 17 MELAB raters, consisted of 13 native speakers of English and 4 non-native English speakers with native-like proficiency, were asked to assess essays written by examinees from 21 language backgrounds. On their own admission, however, this finding is worth questioning, due to several factors. Firstly, all the non-native English speaking (NNS) raters involved in this study have native or native-like proficiency, and the proportion of NNS raters, which was only 24%, is exceptionally small compared to the native English speaking (NES) raters who comprised 76% of the total participants. The second reason is that all these NNS raters live in America (ESL context), where the influence of their first languages and cultures would not be as strong as if they lived in their home countries. Thus, caution should be exercised when generalising to other NNS raters who work in non-English speaking countries (EFL context), such as Indonesia, Thailand and China.

### **2.3.2 Rater types**

In addition to language backgrounds, raters' perceived scoring criterion importance is found to have contributed to the variability of scores in marking. When marking using an analytical rubric, for example, raters may differ in their interpretation of the criteria and these differences would contribute to rater variability. Eckes' (2008) study of 'rater types' and effectiveness of essay marker found that raters were markedly different in their point of views concerning criterion importance. Eckes (2008) clustered raters into four types according to their types: the Syntax type, the Correctness type, the Structure type, and the Fluency type. While the Syntax type

focused on a range of cohesive elements and syntactic structures, the Correctness type was concerned with morphosyntactic, lexical or spelling errors. The Structure type concentrated on the structure of the text, and the Fluency type focused on the degree to which the text could be read fluently (Eckes 2008). The finding, however, contrasted with some earlier work by Hamp-Lyons (1991a; 1991b), which reported the tendency of raters to focus on the syntactic rather than the rhetoric aspects of student essays. Eckes' (2008) findings, regardless of the language background, reverberate with work previously discussed (e.g., Hyland and Anan 2006, H. K. Lee 2009a) in that raters were found to have differences in foci of severity when marking students' work. Eckes' (2012) further work on the rater types showed the correlation between rater's bias pattern and perceived criterion importance in that "criteria perceived as highly important are mostly associated with severe rating while criteria perceived as less importance are mostly associated with lenient rating" (p. 271).

The identification of 'rater types' such as that of Eckes (2008; 2012), is significant in that it raises awareness of the existence of different perceptions of criterion importance among markers and opens the groundwork to further investigate raters' perceptions of the marking criteria used. However, given understanding markers' perceptions of the relative importance of the criteria used, it is worth investigating whether or not these perceptions are reflected in their marking practices. This will provide better understanding of the rater's marking both in terms of reliability of scores as well as the relationship between markers' practices and their perceptions of their practices.

### **2.3.3 Raters' perceptions and practices**

Research into perceptions and practices has become of interest to some researchers, both perceptions of teachers or assessors as well as students or writers (e.g., Brannan and Bleistien 2002, Faez and Valeo 2012, Hyland 2013, H. K. Lee 2008, I. Lee 2009b, Montgomery and Baker 2007). Lee (2008) examined the direct relationship between students' perceptions of a prompt and their actual performance on a field-specific EAP and general topic-writing tests. The study revealed the mismatch between the students' thoughts and performances. A group of 124 students was asked to write two essay samples; one on a regular topic and the other on a field-specific topic. Upon the completion of the writing tests, the students also responded to a survey of their perceptions of the field-specific prompts and tests in general. The students' performances on writing tests were marked by six ESL raters who had undergone extensive training using a holistic rubric. The comparison and analysis of students' performances on the two tests against their survey responses showed that the perceptions were not reflected in the practices. Other related studies (e.g., Ferris 2014, Ferris, Brown, Liu and Eugenia 2011, Hyland 2013, I. Lee 2009b, Matsuda, Saenkhum and Accardi 2013, Pecorari and Shaw 2012) have endeavoured to examine perceptions of teachers and their practices through interviews or questionnaires. The findings from all of the studies, except Hyland (2013), found the differing attitudes and perceptions among teachers in terms of students' needs and feedback. Further, none of the studies systematically investigates how the perceptions were reflected in the work of the teachers (raters) or students (writers), except to some extent, Lee's (2009b) and Hyland's (2013) studies that investigate the correspondence between teachers' beliefs and written feedback practice. The

studies are reviewed below.

Studies into the relationship between teachers' perceptions and their practices have contradictory findings. Lee's (2009b) analysis of EFL teachers' perceptions and their actual practices in a Hong Kong secondary school identifies mismatches between perceptions and practices. The study examined teachers' beliefs and practices in written feedback from two sources: (1) feedback analysis based on 174 texts collected from the teachers and (2) questionnaires and follow-up interviews with them. The study revealed a number of mismatches between the teachers' perceptions and practice, which all can be attributed to teachers' focus on error corrections. The first mismatch is that teachers paid most attention to language form, although they believed there was more to good writing than accuracy. Secondly, teachers tended to correct and locate errors for students but believed that through the teacher's feedback students should learn to correct and locate their own errors. Thirdly, the teachers responded mainly to weaknesses in student writing although they knew that feedback should cover both strengths and weaknesses. Lastly, the teachers continued to focus on student written errors although they know that mistakes will recur. This finding affirms other studies by Hyland and Anan (2006) and Lee (2009a) which shows the tendency of non-native English speaking (NNS) raters (e.g., Japanese and Korean English teachers) to focus more on rule infringement in their marking of students' work. The study of Ferris et al. (2011) of teachers' accounts and feedback also indicates similar results in that teachers pay more attention to ESL students' grammatical errors.

A study by Pecorari and Shaw (2012) provides further evidence of the contradictory

perceptions among teachers. In the study, eight teachers at several Swedish universities were interviewed on their opinions and attitudes towards a set of five pairs of texts extracted from introduction/literature sections of thesis drafts written by two L2 English master's students. The aim was to understand the basis of teacher attitudes and responses to intertextuality in academic writing. The study shows that the teachers were mutually contradictory in their judgements whether or not one text was acceptable, and whether or not it was plagiarised. Although the study did not involve assigning scores, the variability of judgements among lecturers is an indication of lack of agreement or inter-rater reliability, which is of interest to the present study.

Teachers' perceptions and practices are not necessarily contradictory. A study by Hyland (2013) into the correspondence between teachers' perceptions and their actual feedback on students' essays reveals a strong correlation between the two. His study explored the perceptions of feedback to support students in disciplinary writing of teachers working at four faculties at an English-medium university in Hong Kong. These perceptions were gathered through semi-structured interviews. One of the purposes was to uncover the type and frequency of feedback the teachers provided, and what they wanted to achieve with it. The study suggested that teachers' perceptions were concerned more with the rhetorical, genre-level features of argument and disciplinary persuasive logic, rather than grammatical accuracy, and these are reflected in the feedback given (their comments on student assignments). This reflects current approaches to academic writing teaching (e.g., Belcher 2009, de Oliveira and Lan 2014, Hyland 2006, Ravelli and Ellis 2004), which focuses on helping students to mean in ways that are recognised by the

community in academic writing.

This section of the chapter has reviewed issues of rater and reliability in assessment practices. It is apparent from the literature that rater bias accounts for the variability or inconsistencies of scores. There seems to be consensus in the literature on the importance of conducting studies which promote the better understanding of raters' thoughts in relation to their marking practices.

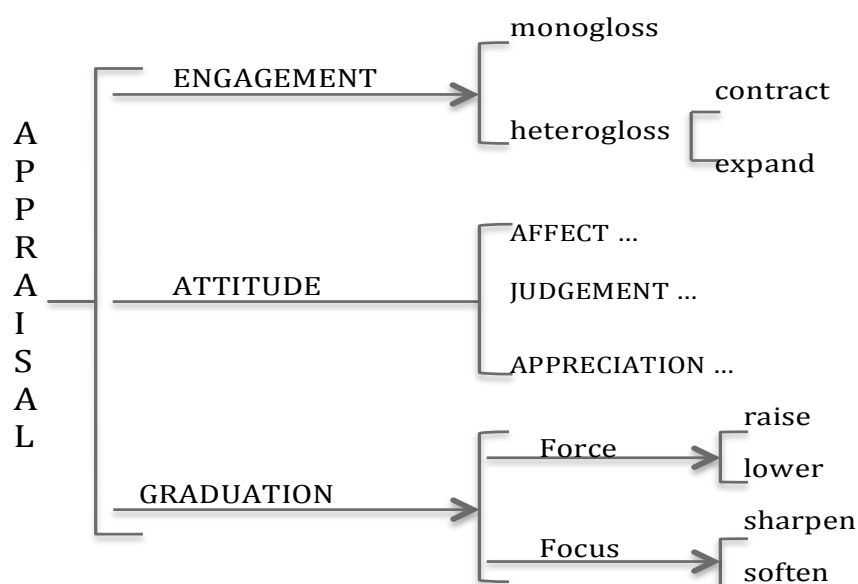
### **2.3.4 Analysis of perceptions using the Appraisal taxonomy**

The Appraisal taxonomy, an extension of SFL, has provided scholars with a powerful tool to analyse voice in texts, and it has been applied in many different fields such as education (e.g., Derewianka 2007, Hood 2004, S. H. Lee 2006; 2010, Wu 2005) and media (e.g., Tran and Thomson 2008, White and Thomson 2008). The emerging theory initially developed through the efforts of a number of linguists, predominantly Martin (2000), White (2002), Martin and Rose (2003) and Martin and White (2005), deals with the Interpersonal dimension of language. A more detailed outline of Appraisal theory is provided in the next chapter devoted to the theoretical framework of the thesis. In this section of the present chapter, a survey of some relevant studies that use Appraisal system to analyse written and spoken texts are presented as they have illuminated ways in which the theory has been applied to such texts. Of the studies, Thomson et al. (2006) Tran (2011) have the greatest relevance for the present study, as they are the only studies that examine the feelings and attitudes reflected in persons' utterances (spoken language) which was gathered through interviews. Past relevant studies by Hood (2004; 2010), Lee, S.H (2006; 2007), Derewianka (2007) and Nakamura (2009) have explored the attitudes

expressed in written texts in educational contexts, but analyses of attitudes in spoken language remain limited. None of the studies, including Thomson et al. (2006) and Tran (2011), directly investigates how the evaluation made is reflected in the teachers' marking practice as the present study is focused on. The studies are reviewed below.

Thomson's (2006) study on The Gardeners' Talk Project reveals the relationship between gardener type and attitudes towards nature and environment. In the study three kinds of gardeners (committed-native, general-native and non-native gardeners) were interviewed on their attitudes and feelings in relation to environmental practices as well as evaluations of behaviour of other people involved in the project. The Appraisal system used to analyse the gardeners' accounts succeeds in teasing out the attitudes of the three interviewed gardeners. For example, committed-native gardener is reported to rely more on Judgement and Appreciation in her responses to the interview questions. The fact that she judges her own behaviour positively, and importantly, would like others to emulate her, shows that she locates herself in a more powerful position than the other two types of gardeners, particularly the non-native gardener who uses Affect the most frequently. While the study certainly paves the way for applying Appraisal taxonomy on analysing perceptions in spoken language, an investigation of Attitude dimension alone is insufficient to effectively provide important insights into the speakers' perceptions in their language use. A more comprehensive study involving the three major system of Appraisal (ATTITUDE, ENGAGEMENT and GRADUATION) would provide a greater depth of understanding. Tran's (2011) study undertakes to do this.

Tran (2011) has successfully used Appraisal taxonomy as analytical tool to deconstruct speakers' perceptions of quality issues of Vietnamese postgraduates. Her unpublished doctoral thesis investigates the attitudes, values and feelings of three groups of people (the managers, the teachers and the students) directly involve in postgraduate education in Vietnam. The three stakeholders were interviewed on their thoughts of the quality of postgraduates in the country and person they held responsible for the problems. The speakers' accounts are analysed using the following model proposed by (Martin and White 2005).



**Figure 2.3 Model of Appraisal resources based on Martin and White (2005)**

The Appraisal system makes it possible to decode evaluation in a text, whether it is spoken or written text, by considering the text in terms of its ENGAGEMENT, ATTITUDE and GRADUATION dimensions, which will be further explored in the next chapter. In the study above, for example, Tran (2011) proves that the Appraisal analysis has been able to demonstrate the relation between the power of the three stakeholders and the dialogic devices found in their talk. In terms of dialogistic

positioning strategy (ENGAGEMENT in Appraisal term), for example, the study has been able to clearly identify the position and power of speakers in making their arguments about quality i.e., whether the speakers are assertive and authoritative, in which case, they would make taken –for-granted statements or they are hesitating and offer space for debate. Briefly saying, as the finding indicates, the more power one has in the system, the more direct is the person’s choice of language in relation to the issues under discussion. In other words, authoritative voice of the participants is depended on their positions in the power hierarchy and whom they talk about. Tran (2011) concludes with the worthwhileness of the intensive and effective examination of texts that Appraisal can bring about for any research that aims at understanding the discourse maker’s feelings, attitudes and values.

In academic setting, studies that involve evaluation in students’ writings have been conducted copiously. The studies are deserving of review in this chapter of the present study because though the studies investigate students’ written texts, they use the Appraisal taxonomy as an analytical tool to deconstruct evaluation in students’ writings in academic setting (e.g., Derewianka 2007, Hood 2004; 2010, S. H. Lee 2006, Nakamura 2009). This will offer some important insights into analysis of evaluative words in a given discourse.

Hood’s (2004) study using the Appraisal taxonomy as an analytical tool to deconstruct student writing in academic setting has been able to offer insights into overall patterns and preferences in the attitudinal choices made in two different kinds of student texts. The study examines how voice or evaluative stance is enacted in the introductory sections of six undergraduates academic writing compared to the

evaluation evident in the same sections of four published research articles. The study uses the Appraisal taxonomy model proposed by Martin and Rose (2003), but focuses on the role of ATTITUDE with resources of implicit (Graduation implicated in evoking ATTITUDE) as well as explicit (inscribed) ATTITUDE.

Hood (2004) classifies all writing as belonging to Field as Domain (FD), referring to information about the content area or the focus of research by the writer and by other source and Field as Research (FR), referring to research itself as an activity. The finding reveals that when evaluating the domain (FD) both student writers and published writers show similar willingness to evaluate explicitly. However, when evaluating field of research activity (FR), they tend to use implicit ATTITUDE, evoked through the resource of Graduation. This suggests that most research appears to be evaluated indirectly (implicitly) by using predominantly the Graduation features.

One major finding related to the present study is that the strong preference across all the texts for ATTITUDE expressed as Appreciation rather than Judgement or Affect. Further, as the study reveals, the published writers' evaluations in the texts are less personalised than those of student writers who also include more expressions of ATTITUDE encoded as Affect. Despite the dominance of Appreciation in both kinds of texts, the student writers' evaluations are predominantly encoded as Appreciation as *reaction* while the published writers evaluate dominantly using Appreciation as *valuation* (construing meaning o significance, usefulness or worth). The preference for Appreciation, as the valuing of phenomena, according to Hood (2004), functions to objectify the valuation to some extent, or at least reflects the nominalised,

objectified nature of the discourse. As *reaction* ‘has to do with the degree to which the text/process in question captures our attention ... and the emotional impact it has on us’ (Martin 2000, p. 160), the dominance of this element of Appreciation in student writers’ evaluations shows a strong connection to their emotional response (Hood 2004). This creates a very much spoken mode of text. In the light of the finding, it is interesting to see whether the ATTITUDE elements sourcing from the emotional response would be more dominant in different discourse such as face-to-face interviews, where the interviewees are required to give direct and spontaneous responses to the interviewer’s questions without having much time to think about what they have to/do not have to say.

A study on the interpersonal development in adolescent writing in the subject History using the Appraisal system has been meticulously done by Derewianka (2007). The study presents a fine-grained analysis of four texts: three were written by early, middle and later years of secondary school students, and the fourth essay by a first year of tertiary education student. The Appraisal system used in the study has been able to reveal how adolescent writers’ use the Interpersonal language to manage social relationships and establishing an authorial identity in their writings. In relation to the evaluative strategies of Attitude, for example, early adolescent student writers are found to rely more on the subsystems of Affect while the more advanced level ones tend to rely on Appreciation and Judgement. With regard to Engagement, Derewianka (2007) concludes that the essays written by students at the lower end of secondary level tend to be undialogised (*monoglossic* in Appraisal term) in that attempts to convince the readers are made primarily through ‘bare assertions’ without showing the needs to negotiate meanings with the readers. The

text written by a first year university student, on the other hand, is richly dialogic (*heteroglossic* in Appraisal term) in that “other voices are explicitly drawn into the discussion, interpreted, analysed, critique and played off against each other” (p. 163).

In a related study, Nakamura (2009) study on the construction of evaluative meanings in English as a second/foreign (ESL/EFL) writing has been able to decode characteristics of high- and low-scored essays in terms of the Interpersonal meanings. In the study, thirteen exam scripts from six proficiency levels (Bands 4 to 9) of the IELTS 9 Band scale were analysed in terms of genre stages following Coffin’s (2004) argument genres and candidates’ evaluations of entities, happenings or state of affairs in their responses to the topics of the task prompts using the Appraisal system based on Martin (2000) and Martin and White (2005; 2007). The detailed Appraisal analyses of student texts taken from the IELTS writing task 2 shows that higher scoring texts, i.e. IELTS band score 7 to 9, are characterised by a preference for *valuation* (Appreciation) over Judgement and Affect. This finding resonates with that of Hood’s (2004) and Derewianka’s (2007).

An earlier study by Lee, S. H (2006) that explores the Interpersonal resources used by East-Asian ESL and Australian-born university students in their argumentative/persuasive essays (APEs) and the extent to which these resources contribute to their relative success in the assessment of their APEs reveals a finding which accords with that of Hood’s (2004). In terms of ATTITUDE resource, for example, the study finds the tendency in the poorer essays to use a more personal or emotional feelings (Affect) rather than an institutionalised feeling using the resource

of Appreciation in the stance taken. The Evaluations in the high-scored essays, on the other hand, are found to predominantly conveyed through the resources of *valuation* and *composition* while a significantly high proportion of Appreciation as *reaction* is found in the poorer essays.

To recapitulate this section, the wide-ranging of literature on the language of evaluation relates to the Interpersonal elements in a text, particularly in an academic literacy field in an ESL and EFL context (e.g., S. H. Lee 2006; 2010, Nakamura 2009, Wu 2005; 2008). There have also been quite a few studies focusing on the application of Appraisal system to different types of discourse. However, up to the present, few studies have been conducted on assessment.

The rich literature has shown the effectiveness of the Appraisal taxonomy, an extension of SFL theory, to examine evaluations and attitudes in texts in various contexts. In educational contexts, for example, the Appraisal system of ATTITUDE, ENGAGEMENT and GRADUATION has been able to features of successful/less-successful texts in terms of the voice and position expressed in the texts. Similarly, in spoken language, the Appraisal analysis has been able to deconstruct the correlation between position and power of speakers and the direct/indirectness of choice of language use in expressing their opinions. This section ends with the call to extent research focusing on the feelings and opinions of teachers as the assessors of students' work.

## **2.4 Summary**

To sum up this chapter, the literature on the assessment shows markers as a

dominant factor in determining the reliability of scores assigned. Despite its importance, there are relatively few studies focusing on the feelings and attitudes of raters involved in the marking processes. This study adds to the relatively small number which have done so (for example Eckes 2008; 2012, Hyland 2013, I. Lee 2009b, Pecorari and Shaw 2012) by providing a case study of an intervention using a linguistically informed rubric and rater training program. It also examines the nature of the markers' experience of the intervention as well as its impact on their practices.

The literature has also shown that there have been some investigations of raters' variables including teachers' perceptions, students' reactions, the importance of training, inconsistency between marking rubrics and actual marking and advantages and disadvantages of rubric types. Despite these, few studies have used a linguistically rigorous tool to investigate those on the whole. For instance, methodological tools employed from previous studies, such as content analysis and thematic analysis (e.g., Ferris, Brown et al. 2011, I. Lee 2009b, Matsuda, Saenkhum et al. 2013, Pecorari and Shaw 2012), does not provide as powerfully systematic linguistic tool as Appraisal analysis does in unravelling teachers' perceptions. The Appraisal taxonomy, an extension of SFL, provides a powerfully systematic linguistic tool for describing teachers' perceptions and attitudes. As we shall see, quality academic essays are also characterised by the choices of Appraisal system.

## **Chapter Three**

### **Systemic Functional Linguistics**

The previous chapter concluded that the explicitness of instructional or assessment practices along with feelings and attitudes of teachers in the assessment practices are areas that need to be explored. The study argues that Systemic Functional Linguistics (SFL) provides a systematic tool for analysing and identifying what constitutes an effective text, and informing the design of explicit assessment marking criteria.

This chapter begins with an explanation of the SFL theory, particularly its key tenets of genre and register, and justification of its use in the study. The usefulness of genre theory in examining teachers' marking practices of students' argumentative essays is discussed. Specifically, in this study emphasis is given to ways in which teachers evaluate the students' texts in terms of stages and phases of argument genres, both discussion and argument essays. Register is considered in terms of how it enables a description of students' linguistic choices regarding the topic of the writing and also of teachers' linguistic choices about the issues raised in the interviews. It also enables a description of the relations between readers and writers (as in student texts) and between speakers and listeners (as in teachers' accounts), their attitudes as well as the relative formality and informality of words used. The last section focuses on the analytical tools informed by SFL, such as Clause Complex, Lexical density/Nominalisation, Theme and Appraisal system that are used in the study.

### 3.1 Language in context

The SFL theory of language offers a systematic description of the relationship between a language and its context of use, positing that meanings are influenced by the social and cultural context in which they are exchanged. There are two contexts that have prominent roles in the systemic theory: the broad context of culture (Martin 1993; 1997, Martin and Rose 2008) and the more specific context of situation (e.g., Halliday 1978, Halliday and Hasan 1985). As will be explicated below (see Section 3.3.2), at the level of cultural context, we can identify the global social purposes of a text. Rose and Martin (2012), for example, identify the various social purposes of a text such as: narrating a complicating event, explaining a sequence of cause and effect, classifying and describing things, arguing for a point of view, and so on. These purposes are realised as **genres** (Derewianka 2012). At the level of situational context, according to Rose and Martin (2012), we can identify the more specific purpose or **register** of a text viewed from its three general functions of **field**, **tenor** and **mode**, which will be explicated in Section 3.3.3.

According to the SFL theory, a text – a complete written or spoken interaction – is not a decontextualised sentence or utterance and it can only be fully understood within its context (Eggins 1994, Lemke 1992). However, due to the interrelationship between language and context, as will be explicated in section 3.3.1 below, it can be argued that there is a strong notion of predictability due to the regularity of patterns in language use (Derewianka 2012). Because of this text and context relationship, we can deduce the context of a text by looking at the text too. Likewise, a text that is likely to be produced in a given context can be predicted because of the social

process that happens over and over again. To put it in a perspective, it is sensible to assume that language choices are influenced by context of situation and context of culture in which a text is produced, and it can only be fully understood when viewed within these contextual perspectives in mind. Regarding the importance of a context, Halliday (1978) argues:

Essentially what this implies is that language comes to life only when functioning in some environment. We do not experience language in isolation - if we did we would not recognize it as language - but always in relation to scenario, some background of persons and actions and event from which the things which are said derive their meaning. This is referred to as 'situation', so language is said to function in 'contexts of situation' and any account of language which fails to build in the situation as an essential ingredient is likely to be artificial and unrewarding (p.29).

Overall, it can be assumed that language can only function naturally when it is used in relation to its context; that is, according to the situational and cultural background of its occurrence. As a result, it is then sensible to say that the language system has also evolved in certain ways so as to meet the needs of its users in different contexts, and can be distinguished from one culture to another. In the present study, for example, students' participation in the language program occurs because of their needs to learn how to use language in writing essays based on the conventions of the English culture of academic writing. Taking Derewianka's (2012) perspective, it can be argued that the language system has certain characteristics due to the evolvement and jobs it does in a certain culture. These characteristics become more distinctive when taken to a different situational context. These contexts are explicated in the next sections.

### **3.1.1 The cultural context of language**

As explicated above, at the level of cultural context, the broad purposes of language

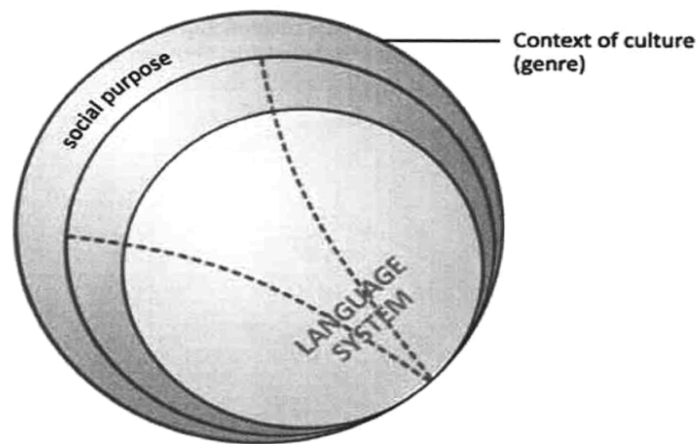
use in society are realised as genres. The present study draws on the ‘Sydney School’<sup>1</sup> genre because of its origin in SFL as a functional linguistic perspective on genre analysis, which provides ways of describing whole texts metafunctionally as well as structurally. This means that the approach of the present study to genre is social semiotic with tenor, field and mode explored as patterned configuration of meaning of social practice (Martin and Rose 2008). Secondly, SFL genre is the theory that many academic writing programs in South-East Asian countries such as Indonesia, Thailand and Vietnam are based on (e.g., Dang 2002, Derewianka 2003, Emilia 2005, Kongpetch 2003, Payaprom 2012, Rose and Martin 2012, Sutojo 1994). In addition, the original rubric used in the school where the data was collected does attempt to account for some aspects of the theory in terms of genre, but as is discussed below, there are limitations to the way it is employed.

As has been explicated above, this study adopts the SFL genre theory. This theory is primarily developed by Martin and Rothery (1980; 1981) and operates at the level of cultural context (see Figure 3.1). Within the SFL tradition, genre is defined as “a staged, goal oriented, purposeful activity in which speakers engage as member of our culture” (Martin 1984, p. 25) or more briefly “staged, goal-oriented social process” (Martin and Rose 2007, p. 8). Genres are referred to as social processes because people participate in genres; goal oriented because genres are used to get things done; and staged because it usually takes more than one step for participants to achieve their goals (Martin and Rose 2007). In other words, the notion of genre is concerned with how a text is organised to achieve its social purpose (Christie and

---

<sup>1</sup> There are three major approaches to genre theory (Hyon 1996): New Rhetoric (e.g., Bazerman 1988, Miller 1984), ESP (e.g., Bhatia 1993, Swales 1990) and SFL-informed genre theory known as ‘the Sydney School’ (e.g., Martin 1985, Martin and Rothery 1980).

Derewianka 2008).



**Figure 3.1 Context of culture (Derewianka 2012, p. 131)**

The explanation above has shown that SFL-informed genre theory is mainly concerned with social practices, in which genres are seen as the recurrent configuration of meanings, which enact social practices of a given culture (Martin and Rose 2008). As a recurrent configuration of meanings, we as members of culture can learn to predict how a situation is likely to unfold, and learn how to interact in it (Martin and Rose 2003; 2007). It can, then, be argued that genres are readily recognisable by members of a particular culture. However, this may not be the case for language users from other cultures. Understanding the use of language patterns of a particular genre is made possible because of our membership in a given culture (Eggins 1994). Consider for example, the stages involved in buying fruit at a market in Australia and in Indonesia. Although some stages in the shopping genre may be similar, others differ. In an Indonesian traditional market culture for example, the prices of things are not fixed and customers are expected to bargain to establish a fair price. Similarly, stages of essay text found in academic writing in English-speaking countries like Australia may have some similarities with those

found in Indonesia. However, they may be different due to the cultural values of these two different societies, which in turn influence their rhetorical structures.

As discussed above, the rhetorical structure used when constructing meaning is highly influenced by the cultural values of a given society. What is valued in one community does not necessarily mean that it has the same value in other communities, especially in two different cultures. In Indonesian culture, for example, an inductive style – providing evidence before giving a legitimate conclusion – tends to be favoured more when reasoning or arguing, particularly when the interlocutor relationship is unequal. Making a claim in the beginning of an interaction can then be considered rude and ineffective. In contrast, other cultures that favour a deductive style, such as people from English-speaking countries, may find inductive ways of reasoning ineffective and “too flowery”. In English academic essays, stating a purpose or claim followed by some evidence or reasons is more favoured. In a very early study, from the perspective of study of rhetoric, Kaplan (1966) using the notion of contrastive rhetoric points out distinctive differences in the written discourse of students from different cultures, one of which described English writing as straightforward, while oriental discourse was symbolised by an inward-pointing spiral. This hypothesis has been discredited by some scholars such as Mohan (1985) and Scollon (1997) among others for its oversimplified generalisation of the rhetorical pattern. These early claims were later revisited and modified. Recently, in response to such crucial critiques, there has been a conceptual shift from oversimplified account in old contrastive rhetoric studies to what is referred to as intercultural rhetoric (Abasi 2012, Connor 2004, S. H. Lee 2013). The variation in writing does not reflect inherent differences in thought pattern but rather

cultural preferences that make greater use of certain options among the linguistic possibilities (Grabe and Kaplan 1996). Thus, understanding the social and cultural context is fundamentally important in making successful language choices, particularly in high stakes academic writing. The present study argues for increasing explicitness to make these culturally-bound conventions more visible to students.

As learning is, overall, a social process, and the environment in which educational learning occurs is that of social institution (Halliday and Hasan 1985), it can be theorised that learning or learners can never be free from the influence of a social institution, society or value systems of a culture.

As culturally formulated activities, genres represent common use of language to achieve particular goals in our society. To accomplish a goal, these formulated activities of language use are organised in terms of series of stages (steps) known as a schematic structure. This schematic structure is “the staged, step-by-step organization of the genre” (Eggins 2004, p. 59) or “the beginning, middle and end structure of texts” (Martin 1985, p. 86). Thus, texts with the same schematic structure belong to the same genre. In a primary school community, for example, Martin (1985) in his early work classifies children’s writing into factual and nonfactual genres. Factual writing “is about the way things are” while nonfactual writing “is about what happened” (p. 5). Included in the factual writing and their purposes are: Procedure (how something is done), Description (presenting specific facts: what some particular thing is like), Report (presenting generic facts: what an entire class of things is like), Explanation (a reason a judgment has been made) and Exposition (arguments about why a thesis has been proposed).

Since then, a number of new genres have emerged as a result of growth in community purposes, as well as in technological and political change (Macken-Horarik 2002). In Rose and Martin's (2012) latest work on genres on secondary school community in Australia, for example, those five main genres for primary school mentioned above have grown into more complex and refined taxonomies. Based on its purpose, for example, the Procedural genre can now be classified into three different types: Procedure (how to do an activity), Protocol (what to do and not to do) and Procedural activity (how an activity was done). Figure 3.2 illustrates the mapping of the genres found in educational settings.



Figure 3.2 Map of genres in schooling (Rose and Martin 2012, p. 128)

As can be seen in Figure 3.2 above, students' texts in the present study belong to the argument genres, in which students may approach to answer the question and persuade their readers to adopt a particular point of view. As we shall see, genres vary according to their purposes and their related staging, and indeed phrasing (see

Figure 3.3 and Figure 3.4 for the stages and phases of argument genres).

The concept of pattern in textual choices valued in English academic writing culture is of interest to this study. This study examines the students' essays in terms of their conformity to stages and phases valued in academic writing culture in English-speaking countries, where the obligatory stages are considered to be a defining feature of a genre (Eggins 2004).

Within SFL, the terms *analytical* and *hortatory* are used to distinguish two different kinds of argument genres. Analytical argument genres are about persuading someone to a particular belief or how the world is. Hortatory argument genres are about persuading someone to do what we believe the world should be (Coffin 2004, Martin 1985). Analytical and hortatory argument genres can be classified further in two categories: one-sided and two- (or more) sided arguments. The one-sided argument is referred to as Exposition genre, while the two- (or more) sided argument is referred to as Discussion genre. In the language school where the study was conducted students' were taught to write analytical Exposition or analytical Discussion essays. Thus, they were expected to convince the readers about their positions on certain issues raised. The classifications of argument genres and their purposes are shown in Table 3.1.

**Table 3.1 A classification of key arguing genres (Coffin 2004, p. 232)**

	<b>one sided argument</b>	<b>two sided argument</b>
<b>'persuading that'</b>	Analytical exposition genre	Analytical discussion genre
<b>'persuading to'</b>	Hortatory exposition genre	Hortatory discussion genre

As has been explicated above, genres, which are organised in terms of schematic structure, are seen as culturally formulated activities of language use to achieve

social purposes. The schematic structures used in arguing genres as seen in Table 3.1, for example, vary in terms of their purposes and strategies used to achieve the purposes. Coffin (2004) summarizes that the purpose of Exposition is to put forward a point of view while the purpose of Discussion is to argue the case for two or more points of view. The relationship between purposes or goals and the phases or stages that the writer moves through to achieve the goals are presented in Table 3.2, with the optional stages shown in brackets.

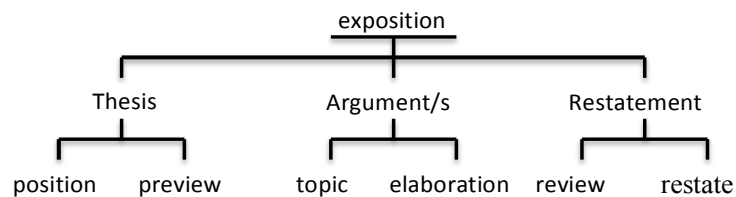
**Table 3.2 Four common arguing genres in student academic writing (Coffin 2004, p. 236)**

Genre	Hortatory Exposition	Analytical Exposition	Hortatory Discussion	Analytical Discussion
<b>Purpose</b>	to put forward a point of view and recommend a course of action	to put forward a point of view or argument	to argue the case for two or more points of view about an issue and recommend a course of action	to argue the case for two or more points of view about an issue and state a position
<b>Staging</b>	<p>Thesis</p> <p>(Recommendation)</p> <p>Arguments + Evidence</p> <p>(Counter-Arguments + Evidence)</p> <p>(Reinforcement of thesis)</p> <p>Recommendation</p>	<p>Thesis</p> <p>Arguments + Evidence</p> <p>(Counter-Arguments + Evidence)</p> <p>Reinforcement of thesis</p>	<p>Issue</p> <p>Arguments + Evidence (2 or more perspectives)</p> <p>(Judgement/Position)</p> <p>Recommendation</p>	<p>Issue</p> <p>Arguments + Evidence (2 or more perspectives)</p> <p>Judgement/Position</p>

As illustrated in Table 3.2 above, it can be seen that the stages of a genre are relatively stable components of its organisation: they are relevant to most instances and can be recognised from one text to another. In analytical exposition and discussion, for example, it can be seen that these kinds of essays consist of obligatory stages of Thesis ^ Argument ^ Reinforcement/Restatement for the former and Issue ^ Arguments (two or more perspectives) ^ Position/Resolution. However, as Martin and Rose (2008) describe, there are more specific phases occurring within each stage, which may vary in terms of their occurrences and sequences. Although feasible, there is no assurance that each stage of a given genre will consist of the

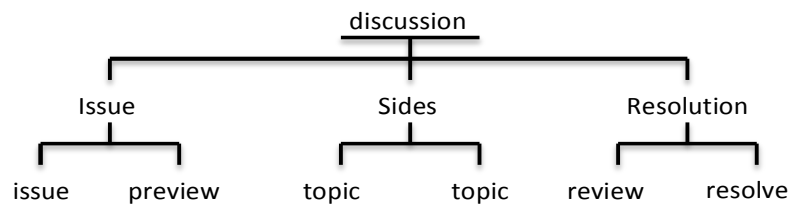
same phases, and even if all of the phases exist, it does not mean that it will happen always in exactly the same sequence. Despite that, as demonstrated by Rose and Martin (2012), some obligatory phases of genres can still be identified.

In this study, I have drawn on Rose and Martin's terminology in describing the phases of the argument genres produced by the Indonesian students and those anticipated by their teachers. Rose and Martin (2012) expand the stages of the argument genres into several phases. The Thesis stage consists of position and preview; the Argument stage consists of topic/hyperTheme and elaboration; and the Restatement stage consists of review and restate. The overall picture is illustrated in Figure 3.3 below.



**Figure 3.3 Stages and phases of exposition (Rose and Martin 2012, p. 272)**

In this study some persuasive texts took the form of discussion. As Rose and Martin (2012) have demonstrated, the discussion phases can be modelled like those of the exposition. The Issue stage consists of preview, the Argument/Side stage consists of topics/hyperThemes, and the Resolution/Position stage consists of review and resolve. The staging of these discussion texts are represented as follows:



**Figure 3.4 Stages and phases of discussion (Rose and Martin, 2012, p. 272)**

So far, this section has discussed cultural context of language. It has explicated the interrelationship between language and cultural context, how context can be deduced from language and stages used and how language features can be predicted by knowing the cultural context. The next section discusses the situational context that influences the language use.

### **3.2.2 The situational context of language**

Language use, whether it is spoken or written, varies according to the type of situation in which it occurs. For example, a conversation with a six-year-old child is not conducted in the same way as a conversation with a sixty-year-old adult. Talk in a lecture is different from a eulogy. Similarly, writing an application letter requires a different kind of language from writing an email or talking face-to-face to a friend. These examples show that language use is influenced by some situational aspects. These situational dimensions can be, but are not limited to, some aspects such as who we are talking to, what is our relationship (close/distant) with the person we are talking/writing to, and what means of communication is used. Halliday (1978, 1985) proposes three situational categories using the notion of register for the contexts of situation. Register is defined by Halliday (1985) as:

a configuration of meanings that are typically associated with a particular situational configuration of field, mode and tenor. But since, it is a configuration of meanings, a

register must also, of course, include the expressions, the lexico-grammatical and phonological features, that typically accompany or release meanings (pp. 38-39).

The context of situation consists of three variables (field, tenor and mode) known as the register variables (see Figure 3.2). Halliday (1985) describes these three situational categories as follows:

- (i) **Field** refers to what is happening, to the nature of the social action that is taking place: what it is that participants are engaged in, in which language figures as some essential component.
- (ii) **Tenor** refers to who is taking part, to the nature of participants, their statuses and roles; what kinds of role relationships exist, including permanent and temporary relationships of one kind or another, both the types of speech roles they are taking on in the dialogue and the whole cluster of socially significant relationships in which they are involved.
- (iii) **Mode** refers to what part of language is playing, what it is that the participants are expecting language to do for them in the situation: the symbolic organisation of the text, the status that it has, and its function in the context (Halliday and Hasan 1985).

Briefly, these three register variables, which influence language choices in a situation, can be summarised as: the “what” or “the subject matter” (field), the “how” or “the means of interaction” (mode), and the “who” or “role relationship between the interactants” (tenor).

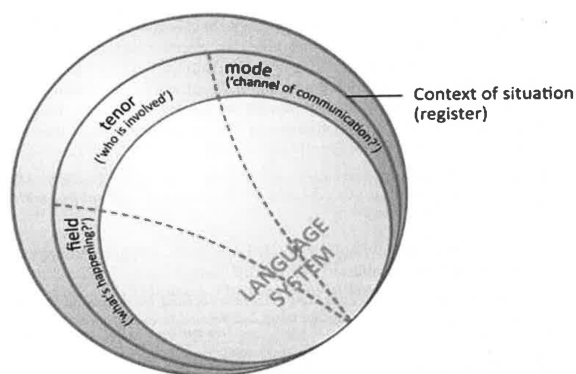


Figure 3.5 Context of situation: register (field, tenor and mode) (Derewianka 2012, p. 132)

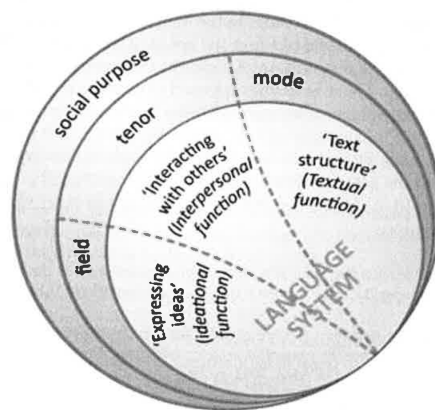
In the present study, field refers to the topics or issues that arise from the teaching and learning of academic English in the language school (college) setting. These may include issues that students have to address in their essays and issues relating to

teaching and assessing practices.

In terms of tenor, there are two kinds of relationship involved that influence interpersonal choices made by both the students and the teachers. Firstly, there is an unequal relationship between the students and the task given as the representation of the language institution and its values. In this case, the language institution where the students are studying is a powerful social institution that has more power to demand from students. The second strand is the relationship between the researcher/interviewer and the teachers/interviewees. The second relationship is the unequal, more superior power between teachers and students. The fact that teachers have power over assessment is the key factor in this unequal relationship. These vertical and horizontal relationships will likely influence the interpersonal choices they make when expressing their feelings and attitudes.

Mode refers to various uses of language in the language school (college) context. For example, the compact, carefully crafted language of written texts, such as in student essays, will be different from the spontaneous language of spoken texts, as in teachers' accounts.

There is a one-to-one correlation between the three variables of the context of situation (field, tenor and mode) and the three metafunctions discussed earlier. The field of a discourse relates to the ideational meanings; the tenor of the discourse relates to the interpersonal meanings; and the mode of the discourse relates to the textual meanings. The correlation between context and metafunction is illustrated in Figure 3.6.



**Figure 3.6 Register variables and metafunctions correlation (Derewianka 2012, p. 137)**

In the present study, the significance of register is two-fold. The first significance is related to students' work in what is a new experience for them. When Indonesian students take an academic writing program in an English as a Foreign Language (EFL) context, they have to adjust to a new context of situation to be able to write argumentative essays. This is because they have not written in these genres in previous settings. In terms of field, students need to develop control over discipline knowledge such as globalisation and its effects on developing countries. With respect to tenor, they are requested to take on new roles so as to enable them to engage and communicate their opinions in an appropriate way as required by the language institution they study in. In terms of mode, they need to learn how to manage their message in an academically acceptable written language and organise it in such a way so that the meaning is clear to the readers.

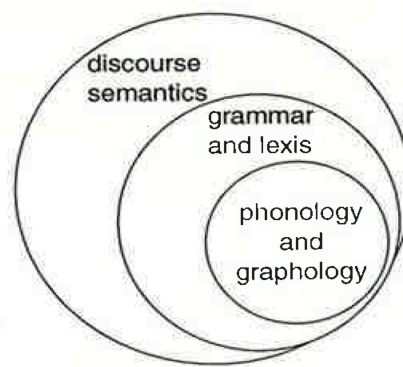
Register is also significant as an organising principle for the marking rubric design for use in the intervention. The register variables of field, tenor and mode helped inform the design of the marking rubric to assess more precisely students' use of the three metafunctions in their essays. In doing so, the marking rubric developed was more thorough as it covers choices at the level of genre and register as they are

realised in language.

In addition, the present study also analyses teachers' ways of expressing their attitudes and feelings and how they engage with their listeners. For this purpose, the interpersonal meaning, as part of register theory, is deemed significant for this part of the study. This evaluative language is explicated in the following section using the term Appraisal, which was developed by Jim Martin and colleagues (e.g., Martin 2000, Martin and Rose 2003; 2007, Martin and White 2005, White 2002).

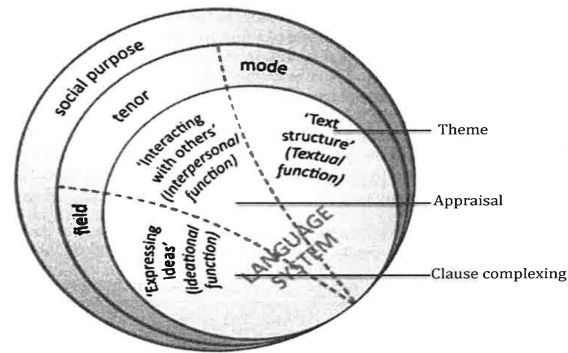
### **3.2.3 Language strata and systems**

Within SFL, there are two major important notions in considering language: metafunction and realisation (Halliday 1978). As has been explained above (see Section 3.2), metafunction is the idea that language serves three different kinds of meanings namely: the ideational, the interpersonal and the textual functions. In terms of realisation, language is recognised as a system comprising three strata (see Figure 3.7). The outermost stratum, which is concerned with meanings at the level of the text and paragraph (beyond the clause), is referred to as the discourse semantics stratum (Martin 1992, Martin and White 2005). These meanings are realised by the choices made from the wording stratum, referred to as lexico-grammar – a combination of 'lexis' (vocabulary) and grammar (Derewianka and Jones 2012, Rose and Martin 2012). This is expressed through the stratum of phonology (as in spoken language) or graphology (as in written language). Beyond the discourse semantics, there are extralinguistics levels: the previously mentioned context of situation and context of culture (see Section 3.3.1 and Section 3.3.2 above).



**Figure 3.7 Language strata (Martin and White 2005, p. 9)**

As previously discussed, SFL theory is concerned with language as a system of resources for making meaning in relation to the context where it occurs. In making sense of the meanings the two contexts discussed above (genre and register) use language systems as their expression planes. As relevant to this study, these systems of meanings can be classified into: the Clause Complex, Appraisal, and Theme systems. As Halliday (1985a; 1994) explains, the clause complex system enables us to account in full for the functional organisation of sentences. Clause complexing is concerned with the logical relation of one clause to another. As a clause consists of information about people, processes, things, the system of clause complexing belongs to the ideational function of language and is related to the field of the text. The Appraisal system is concerned with words used to evaluate feelings, people and things (Eggins 2004). Since it is concerned with evaluating feelings and power relationship in interactions, its function is interpersonal and is related to the tenor in a text. The Theme system is concerned with the organisation of the point of departure of a message in clauses and texts. Thus, its function is textual and is related to the mode of the text. The overall picture of the relationship between language systems, functions, and their contexts is displayed on Figure 3.8.



**Figure 3.8 The context relationship with the language functions and systems (Derewianka and Jones 2012, p. 37)**

As illustrated above, the context of situation (register) is seen as the immediate environment of language, while context of culture (genre) is seen as a broader background against which the text has to be interpreted. The relationship between register and genre perspectives is treated as an inter-stratal one, with register realising genre, and can be seen as recurrent configurations of register variables of field, tenor and mode. The relationship between register and genre, in other words, is treated as similar to that between language and context. Thus, genre is seen as a pattern of register patterns, just as register variables are patterns of linguistic ones.

### **3.2.4 The clause complex system**

The clause complex system belongs to the *logical* metafunction, which is part of the *ideational* metafunction of language, and closely related to the register variable of field in the context of situation (Halliday 1978; 1985a). The ideational metafunction concerns how people construe experiences to make sense of the world and the logical relation between the experiences in clauses (Matthiessen 1992). It should be noted that there is another way of approaching the Systemic Functional Linguistics i.e., Martin's (1992) four discourse semantics (Ideation, Conjunction,

Negotiation/Appraisal and Identification). Of these two divergent houses, I am using the logicosemantic system from grammar to do analysis of *ideational* function of language, but I am drawing on the discourse semantics system of Appraisal to analyse the *interpersonal* aspects of students' essays and teachers' talk.

Clause complexing occurs when two or more clauses are linked together. Clause complexing formation involves two kinds of sub-systems: the tactic, which describes the interdependency relationships between clauses, and the logico-semantic system, which describes the specific types of meaning relationships (Halliday 1985a, Halliday and Matthiessen 2004). Of particular interest to the study is the taxis system, as this is where the organisation of experiential meanings can be observed. In this system, the options of interdependency relationship can be realised in equal, independent relationship (parataxis) or unequal status with dependency to a main clause (hypotaxis).

#### **3.2.4.1 Lexical density and Nominalisation**

Written texts are more lexically dense than the spoken ones (Halliday 1985b, Martin 1985, Ravelli 1988). Halliday (1985b) proposes that both spoken and written language are complex in their own way. Spoken language is organised around the clause i.e., the information is typically organised more intricately, with more clauses to accommodate the lexical material. By contrast, the typical kind of complexity in written language is lexical density i.e., the information is typically presented with “a much denser pattern of lexicalised content” (Halliday 2001, p. 182).

The high degree of lexical density in written texts is made possible by the use of grammatical metaphor, in that the extra information is packed into the participants to create long nominal groups. In fact, mature writing is characterised by some metaphorical mode of expression (Halliday 1985a; 1998, Halliday and Matthiessen 1999, Ventola 1996). Students' control over this mode of expression is central to success in the higher levels of schooling – the construal of experience into specialised domains and the reasoning about experience in abstract, logically developed terms (Christie and Derewianka 2008, p. 25).

One of the key reasons for lexical density is nominalisation, in that highly nominalized written text allows a far greater proportion of the words in the text to be 'content carrying' (Eggins 1994; 2004, Ravelli 1988). Further, nominalisation is a feature of mature writing in that it is essential for dealing with abstract knowledge (Halliday 1985b, Knapp and Watkins 1994). Because written language usually creates an entire context of something, it must be self-contained, it must name entities and processes. Therefore, it is more lexically dense than spoken language which can usually refer to entities in the immediate context by using pronouns and gestures. Thus, in argument essay writing, certain characteristics of the written mode, such as nominalisation, high lexical density and low grammatical intricacy are expected.

### **3.2.5 Theme**

The system of Theme belongs to the textual metafunction of the language, and it is most closely related to the register variable of Mode in the context of situation. This system is concerned with "the organisation of information within clauses, and

through this, with the organisation of a larger text” (Martin, Matthiessen and Painter 1997, p. 21). This system is realised through a structure in which the clause falls into two main constituents: a *Theme* and a *Rheme*. A Theme is a point of departure of a message telling what the clause is going to be about (Eggins 2004, Halliday 1994, Halliday and Matthiessen 2004, Martin and Rose 2007). A Rheme is the part of the clause in which the Theme is developed, and typically contains new information (Eggins 2004, Halliday 1985a). In other words, a Rheme signposts “ the temporary destination”, telling the audience where they are heading (Butt, Fahey, Feez, Spinks and Yallop 2000, p. 142). Simply put, Theme refers to what is being talked about (the subject), whereas Rheme typically refers to what is being said about the Theme.

### **3.2.6 Appraisal theory**

Concerned more particularly with the language of evaluation, attitude and emotion, Appraisal theory (Martin 1997, Martin and Rose 2007, Martin and White 2005, White 2002) is used as an analytical framework in investigating the interpersonal dimension of teachers’ use of evaluative language in the present study. As a semantic tool of analysis of evaluation by speakers/writers, Appraisal taxonomy provides a systematic tool to help with exploring, describing and explaining the way language is used to evaluate, adopt stances and to manage interpersonal positioning and relationships (White 2001; 2012). The theory has emerged from within Systemic Functional Linguistics (see, for example Halliday 1985a, Halliday 1994, Halliday and Hasan 1985, Martin 1992) and is concerned with words used to evaluate feelings, people and things. This theory originates from a project conducted in the 1980s and 1990s as part of the Disadvantaged Schools Program’s *Write it Right* literacy project (Rose and Martin 2012).

Within SFL, Appraisal can be located as an interpersonal system at the level of discourse semantics, and can be regionalized into three interacting domains – ATTITUDE, ENGAGEMENT and GRADUATION. ATTITUDE is concerned with our feelings, including emotional reactions, judgements of behaviour and evaluation of things or phenomena. ATTITUDE can be divided into three categories: Affect (expressions of a person's feelings), Judgement (expressing normal judgement of people's behaviour), and Appreciation (expressing aesthetic assessments of objects valued by society such as artworks, states of affairs or people (but not their behaviours)). ENGAGEMENT deals with sourcing attitudes and the play of voices around opinions in discourse. The voices according to the ENGAGEMENT can be grouped into Monogloss and Heterogloss. According to White (2012), monoglossic attitudes are those that have no recognition of other voices and are about the speakers' commitment to the truth-value of their utterances. Heteroglossic attitudes, on the other hand, recognise other voices and allow the speakers to characterise their utterances as less than factual or less than certain knowledge. GRADUATION attends to grading phenomena whereby feelings are amplified and categories blurred (Martin & White, 2005). The Appraisal system makes it possible to decode evaluation in a text, whether it is spoken or written, by considering the text in terms of its ATTITUDE, ENGAGEMENT and GRADUATION dimensions.

In the context of the present study, a close examination of the ATTITUDE, ENGAGEMENT and GRADUATION elements provides insights into the opinions and values of the teachers in relation to assessment practices and procedures and of the students in terms of the topic of the assignment given.

When students write their responses regarding the topic given, they are expressing their attitudes towards the topic of the study and negotiating their positions. At the same time, they are also including and excluding other sources of voice in their writing while showing their appreciation of the phenomenon or making judgement of the people involved in it. Appraisal describes the linguistic resources for doing so and is therefore useful in designing explicit teaching and assessment procedures. Similarly, when teachers are responding to students' writing, they are making intersubjective judgement as well. The Appraisal system provides tools to assist me to 'get at' the teachers' attitudes, because they are engaging and appreciating students' exposition texts and judging students as a result. Therefore, the Appraisal system, as part of the interpersonal metafunction, will be used as a tool to analyse participants' opinions and attitudes, which are gathered through face-to-face, semi-structured interviews.

Research on Appraisal (see, for example, Derewianka 2007, Hood 2004, S. H. Lee 2006; 2007, Nakamura 2009) has revealed that quality academic essays are characterised by high number of more objective Appraisal choices (Appreciation) rather than more subjectively oriented direct expressions of feelings (Affect) or opinions of people's characters (Judgement). Therefore, the Appraisal system will also be used as one of the criteria in the assessment of student's essays.

ATTITUDE (Affect, Appreciation and Judgement) as a system of meaning for mapping feelings can help me tease out teachers' feelings in terms of emotion (Affect), aesthetics (Appreciation) and ethics (Judgement). Teasing out these is deemed important because the present study is concerned with attitudes and

reactions to students' work and to assessment processes. Furthermore, an understanding of the teachers' positive/negative attitudes and reactions will aid the evaluation of the intervention program.

ENGAGEMENT resources are helpful in understanding students' and teachers' value position. When talking about feelings, reactions or assessments, for example, the teachers not only self-expressively 'speak their own mind' but simultaneously invite others to endorse and to share with them the feelings or assessments they are announcing (Martin and White 2005, White 2012). In other words, the teachers positioning can be analysed in terms of their *monoglossic* and *heteroglossic* utterances. The present study is concerned with analysing teachers' attitudes and perceptions of (1) the marking rubric used, (2) their consistency in marking student essays, and (3) the intervention or rater-training program given. Thus, analysing teachers' *monoglossic* and *heteroglossic* utterances is deemed important because such analysis can further inform me as to whether the speakers' positions or viewpoints are firmed or whether they are opened to other alternatives.

GRADUATION resources are helpful in this study because teachers' assertions when expressing their attitudinal meanings and engagement values are made gradable. In other words, the degree of teachers' negative and positive evaluations of their reliability in marking, the helpfulness of the marking rubric used, their perceived importance of the criteria in the rubric used and other aspects, as of interest to the present study, can be made more gradable. The systematic analyses of teachers' perceptions using the three Appraisal resources (ATTITUDE, ENGAGEMENT and GRADUATION) will certainly offer insights into the language use. Some samples of

how the Appraisal resources are used to analyse the teachers' opinions will be explicated in the next chapter (Chapter 4, Methodology).

### **3.3. Summary**

This chapter has explained the theoretical approach employed in this study – Systemic Functional Linguistics (SFL) theory. The cultural and situational contexts of the given theory (genre and register) inform the design and development of the intervention. More specifically, with respect to the cultural context, the marking rubric will articulate the language choices in terms of the selection and staging of genre to achieve the assessment task. The register variables of field, tenor and mode inform the development of criteria, which articulate the language choices at the level of text and clause in terms of topic and content (field), formality and stance (tenor) and written features to achieve cohesion (mode). The Appraisal system, serving as the interpersonal function of the SFL theory, provides a systematic linguistic tool to analyse opinions, feelings and positions the students and teachers make and at the same time, this system can be used as one of the criteria to assess the quality of students' essays.

## **Chapter Four**

### **Methodology**

This chapter describes the methodology, including the inquiry processes used in the study. The previous chapter outlined the theoretical framework for the thesis. It concluded with the significant contribution of SFL theory (Halliday 1978; 1985a) to inform the design and development of the intervention program. It was argued that the use of a marking rubric based on the coherent model of language offered by SFL would lead to better assessment practices. The contextually oriented concepts of genre and register in conjunction with the linguistic system of Appraisal were found useful in describing the quality of the teachers' assessment practices and their attitudes and beliefs about these. Genre informed the analysis of revised assessment practices and advocated the intervention in terms of staging and phasing used to achieve the purpose of the assessment task. The register variables informed the more explicit assessment criteria in terms of how the students managed the content, how they managed the reader-writer relationship and how they organised their texts. The Appraisal system provides a systematic tool to analyse speakers' or writers' evaluations as language choices. This chapter begins with an account of the research purpose and the design to achieve it. Then it continues with some information about the participants, data collection processes and quality of the research. The final section of the chapter details the analytical devices used for the study as informed by the theoretical framework.

## 4.1 Research purpose

As has been established in the previous chapter, the main purposes of this study are to identify the nature of the problems that exist in the current assessment of student written texts in an Indonesian context and to determine the efficacy of a more linguistically explicit marking rubric. In particular, the study seeks to examine the effect of marking criteria on inter-rater reliability. For this purpose, the participants were required to attend a four-session rater-training program, at which they were introduced to some aspects of Systemic Functional Linguistics (SFL) theory and an SFL-informed marking criteria. The impact of the intervention on assessors' marking practices in terms of their perceptions of consistency and the training given was then explored. Table 4.1 below provides the links between research questions and the methods of collecting data and tools of analysis.

**Table 4. 1 The links between purposes/research questions and methods**

Purposes/research questions	Phases	Instruments for data collection	Data analysis tools/methods
to examine Indonesian EAP teachers' essay marking experiences using the existing marking rubric and to identify the most problematic aspects of student essays to mark	<b>one</b>	interviews  essay marking activities (10 essays)	Appraisal analysis  consensus estimate of inter-rater reliability
to introduce teachers to the SFL-informed rubric	<b>two</b>	intervention/rater training	SFL theory  concepts of genre, register and the systems of Clause Complex, Theme and Appraisal  practices on analyzing sample essays using the SFL-informed rubric
to identify the effects of SFL-informed rubric on teachers' marking consistency and teachers' responses to the intervention	<b>three</b>	essay marking activities using SFL-informed criteria (10 essays)  interviews	consensus estimate of inter-rater reliability  Appraisal analysis

## **4.2 Research design**

### **4.2.1 Qualitative approach**

A qualitative approach was employed to address the research questions posed in the previous chapter as the study adopted a number of characteristics typical of qualitative research. These included aims to understand participants' perspectives, categorical evidence dependence, the human research instrument, rich descriptions and naturalistic setting. First of all, as the researcher, I was concerned with my participants' perspectives or personal understandings of the issues (Arksey and Knight 1999, Bogdan and Biklen 2003): what assumptions they made about the marking rubrics, their rating practices and the intervention program. In other words, this study relied on categorical evidence such as perceptions or opinions rather than numerical evidence in answering the research questions (Richards 2003, Yin 2003; 2009). Secondly, qualitative research is best suited for research problems in which unknown variables need to be explored (L. Cohen, Manion and Morrison 2000). Thus, in the present study, some explorations on what caused problems in teachers' marking practices were necessary prior to designing the intervention program. Further, as the literature yielded little information about the phenomenon of my study, i.e., the relationship between teachers' thoughts and beliefs with their assessment practices, I wanted to learn more from the participants through exploration of their perceptions and practices (Creswell 2005).

A qualitative approach to inquiry is also characterised by the collection of data in a natural setting, that is, in the field at the site where participants experience the issue or problem under study (Bogdan and Biklen 2003, Merriam 1998, Richards 2003).

In this study, the issue or problem was related to the teachers' assessment of student essays in an English for Academic Purposes (EAP) program at a language institution in Indonesia. The rating processes in this study were not conducted in a laboratory or a special class. Instead, they were conducted at the site where the participants usually marked their student essays (teachers' offices and homes), with as little disruption to the natural setting as possible (Creswell 2007; 2009, Merriam 1998). Further, major data source were essays students prepared and wrote in their homes and submitted in the course of their regular classroom activity.

#### **4.2.2 An action-oriented case study**

The study is sought to examine the Indonesian English teachers' marking experiences at a single language school at a particular time, using an action-oriented single-case study. Action-oriented research aims to achieve a profound understanding of the performance of people in real world (Pihlanto 1994) and to improve practice or develop individuals' capacities or skills (Herr and Anderson 2005). In this study, the action was done through an intervention program, designed to inform teachers' performance in essay marking. A single case study can be defined as a study that is conducted using just one incidence or example of the case or phenomenon at a single site (Abercombie, Hill and Turner 2006, Babbie 2004, Lapan, Quartaroli and Riemer 2012). In the present study, the single case comprises one language school, a particular EAP assignment and one group of teachers who regularly engage in similar marking activities. Case studies focus on the in-depth exploration of a program, event, activity, process, or one or more individuals, and are bounded by time and activity (Babbie 2004, Creswell 2007; 2009, Kumar 1996, Lapan, Quartaroli et al. 2012, Stake 1995). A case study is considered appropriate

for this research because it is the most suitable means of exploring and explaining the ‘what’ and the ‘how’ of the questions (Yin 1994; 2003; 2009) or to unearth new and deeper understanding of phenomena (Lapan, Quartaroli et al. 2012).

Based on its special features, a qualitative case study can be characterised as being particularistic, descriptive and heuristic (Merriam 1998 p. 29). This study was particularistic because it focused on a particular situation, event, program or phenomenon, i.e., it focused on teachers’ marking of EAP students’ essays in a particular language institution in Medan City, Indonesia. It was descriptive because its end product was a rich, thick description of the phenomenon under study, i.e., it relied on linguistically informed analyses and interpretations. A heuristic quality of a case study can explain the reason for a problem, the background of a situation, what happened and why, and explain why an innovation worked or failed to work (Merriam 1998). In this study, the case was a group of EAP teachers. Specifically, the focus of the study was the marking practices of a group of teachers.

#### **4.2.3 Participant selection**

The site for data collection was an English language institution in Medan, the capital city of North Sumatra province, Indonesia. One of the courses provided by this language institution is English for Academic Purposes (EAP) course. The EAP course in this institution comprises six levels: EAP 1–EAP 6. This study particularly focused on EAP 4 because at this level students are asked to write two essays in argument genres (see, for example, Coffin 2004, Martin 2005, Martin and Rose 2008): a single perspective exposition and a multiple perspectives discussion, as part of requirements to pass the course. The study explored EAP teachers’ marking

students' argumentative essays at the research site.

Six English as a Foreign Language (EFL) teachers, identified as Dina, Jelita, Mirzan, Ratna, Yusuf and Whitney, were recruited by using a *convenience sampling* method. In this sampling method, I did not choose a random sample that could represent the whole population, but selected a sample that suited the purposes of the study and that was convenient (Gall, Gall and Borg 2007, Merriam 1998). The sample was convenient for a variety of reasons: the site was familiar to me thus avoiding extra time establishing relationships, and I could have immediate impact on policy and practices at the institution. The sample was located at the institution where I worked, and I was very familiar with the site and the participants, and thus could bring additional insights.

In order to exclude the effect of different teaching experiences on the variability in teachers' rating performance (such as in Barkaoui 2010a, Weigle 1994; 1998; 1999), the participants were recruited using the *purposive sampling* method. As the name suggests, participants were deliberately chosen because they had "special characteristics or experiences that can contribute to an understanding of the phenomenon studied" (Hennink, Hutter and Bailey 2011 p. 84). The participants for this study were chosen because they had a minimum of two-years experience in teaching and marking EAP or academic writing at the language school where the research was conducted. There were six EAP teachers recruited, and they all participated voluntarily in this study. Four of the teachers were female and two were male, and all of them had had four to six years experience teaching EAP, International English Language Testing System (IELTS) and academic writing. Five

of them were native-speakers of Indonesian and one was a native-speaker of English, who had lived in Indonesia for approximately four years and could communicate fluently in Indonesian.

Initially, only native Indonesian speaking assessors were intended to be recruited as participants in this study. However, as the results of the previous research (e.g. Hamp-Lyons and Davies 2008, Hyland and Anan 2006, Johnson and Lim 2009, Lim 2009, Shi 2001) were varied in terms of the influence of assessor's language background to rating, I decided to include one native English speaking assessor (identified as Whitney) as a participant in this study. In addition to the results of previous research, another reason to include Whitney was that she was a member of a group of teachers who worked together in the same language school, and that program being the case in the study, I wanted to keep the team that worked in the program together. In addition, she spoke Indonesian fluently, was married to an Indonesian, and understood the culture well. All five other teachers were fluent in English. As I was not focusing on the comparison of native and non-native English speaking assessors, it was therefore appropriate to include Whitney in the study. The information about the research participants is presented below (see table 4.2).

**Table 4.2 The research participants**

No	Name	Gender	Age	Experience	Language background
1	Dina	Female	35-40	6 years	Indonesian
2	Jelita	Female	25-30	4 years	Indonesian
3	Mirzan	Male	35-40	6 years	Indonesian
4	Ratna	Female	35-40	4 years	Indonesian
5	Yusuf	Male	30-35	4 years	Indonesian
6	Whitney	Female	40-45	4 years	English
Note: Names are pseudonyms.					

Note that for ethical reasons it is not possible to mention the exact ages of the participants here or anywhere else in the thesis. The number of teachers working at the language institution, particularly those whose first language is English, is relatively small. Thus, mentioning their exact ages may alert readers to the identities of the participants.

### 4.3 Data collection

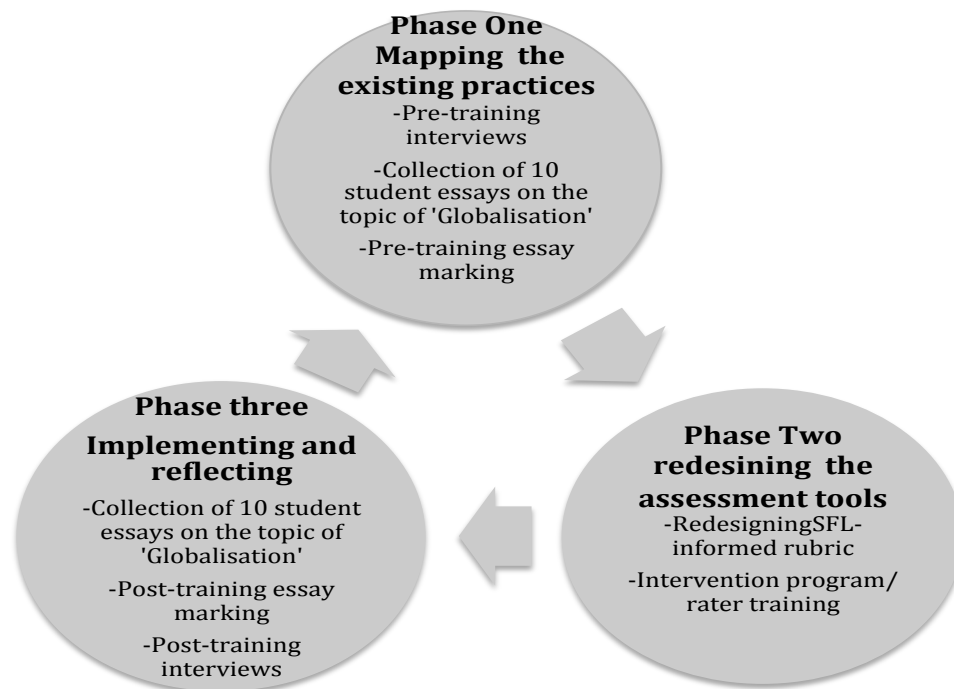


Figure 4.1 Phases of data collection

As shown in Figure 4.1 above, the data collection process lasted approximately three months and comprised three phases. The first phase involved collecting information about the six EAP teachers' assessment practices in the language school through interviews and observation of marking practices. As seen above, the interviews were conducted in two phases: pre- and post-training. All of the interviews were transcribed verbatim for analysis. Transcribing was done soon after each participant

had been interviewed (Kvale 1996), and the transcripts of the interviews were given to the participants so that they could check the correctness of the transcriptions on their accounts (Stake 1995; 2010). Since all of the participants had work commitments, they were given one week to read, comment and return the transcripts. Although transcribing the interviews was quite time consuming, it had assisted me to become very familiar with the data. The multiple marking of the sample essays was on the same topic: “The Effects of Globalisation on Developing Countries”. Twenty essays were collected: ten for pre-training marking and ten for post-training marking. The pre-training marking used the existing rubric provided by the language school (see Table 4.3).

**Table 4.3 The existing marking rubric**

<b>EAP-Plus-4 Assignment 2 (Argument Essay)</b>					
<b>General Statement</b>	1	2	3	4	5
<b>Thesis Statement</b>	1	2	3	4	5
<b>Topic Sentences</b>	1	2	3	4	5
<b>Supporting Sentences</b>	1	2	3	4	5
<b>Subject-Verb Agreement</b>	1	2	3	4	5
<b>Vocabulary, Spelling, and Punctuation</b>	1	2	3	4	5
<b>Conclusion</b>	1	2	3	4	5
<b>Cohesion, Readability and Relevance</b>	1	2	3	4	5
<b>Overall grade</b>					
<b>Length</b> <input type="checkbox"/> Underlength <input type="checkbox"/> (minus 1 point)					
<b>Final grade (Divided by 1.33)</b>					
Comments: _____					

After analysing the data gained from Phase One, a linguistic-based intervention program was designed with the purpose of making more visible and therefore more explicit to the teachers the linguistic markers of quality academic essays so that there would be more agreement in their marking practices. In the last phase,

information about assessment practices using the new marking rubric was collected by interviewing the six teachers and analysing their marking consistencies.

The steps in Phase One and Phase Three were conducted in reverse order, that is, in Phase One interviews were followed by essay marking activities while in Phase Three essay marking activities were followed by interviews. The differences in the sequence of the data collection were determined by my understanding of the participants' experiences with the marking rubrics used. Phase one started with interviews in order to explore teachers' perspectives on their essay marking consistency and of the existing rubric used to mark the essays. As the participants had worked in the same language school for more than two years and the information needed was related to their routine marking practice, they were all well placed to respond to the topics of the interviews. Therefore, it was logical to interview them first before they marked the student essays. In contrast, in Phase three the teachers had just finished training in SFL and in using the SFL-informed rubric, and had not yet had enough exposure to the new rubric. Thus, they needed experience at using the rubric first before being able to give opinions about it and its impact on their consistency in marking.

#### **4.3.1 Phase one – Exploring exiting practices**

Phase one addressed Research Question 1 (hereafter will be called RQ1) *How do raters perceive the existing marking criteria?* and RQ2. *What are the main sources leading to rater inconsistency in student essays?* To answer the two questions, one-to-one, semi-structured teacher interviews were conducted. The interview technique was used on the grounds that it enabled me as the researcher to collect in-depth

information by engaging with each interviewee's perspectives (Patton 2002). Semi structured interviews gave an element of structure to the interview process, and at the same time allowed flexibility to explore areas of interest, or to examine areas that I might not have foreseen prior to the interview. The participating teachers were then asked to assess ten student essays using the marking rubric already in use in the language school. Informed by Stemler's (2004) formula, the teachers' marking practices were analysed in terms of the similarity of scores given to each of the essays (inter-rater reliability) and features of the essays with the most and least agreement. These features would be relevant to designing the intervention.

During Phase One the teachers were asked questions such as: "Do you find difficulties in marking your student essays?"; "What features of the essays do you find the most difficult to mark?" or "If you could add some additional detail to this criteria, what would you add?" and others (see Appendix C for the interview guide). The purpose of these questions was to identify and then explore teachers' perceptions of difficulties with current practice using the existing, non-SFL marking rubric.

#### **4.3.2 Phase two – Redesigning practice**

##### **Design of the training program**

SFL theory informed the development of the training program. As explicated previously, one of the main purposes of the study was to enhance consistency in marking by introducing teachers to SFL descriptions of the academic essay, and the intervention and the rubric helped this. SFL theory posits that language and its contexts of use are systematically related. The two contexts that have prominent

roles in the systemic theory are the broad context of culture and the more specific context of situation. In other words, language use generally can be identified based on its social purpose (known as genre), and based on the situations in which it takes place (known as register). Therefore, it was sensible to start the training program with the concept of language functions and concepts of contexts. Moreover, SFL was relatively a new entity in this teaching context, and only one participant had encountered it previously. Therefore, it was necessary to introduce and provide general overview of the theory and its key constructs before introducing, explaining and practising with the SFL-informed rubric.

Understanding the concepts of genre and register would be beneficial for the teachers' marking practices. By understanding the concept of genre, for example, the teachers would be able to assess students' essays based on how effectively they met the social purpose of the task, and the stages and phases of the essays through which the purpose is achieved. In this study, for example, the purpose of the essays was to argue for or against the impacts of globalisation on developing countries. Similarly, the teachers' understanding of the concept of register would help them rate the appropriateness of an essay in terms of the writer's control over discipline knowledge or relevant issues, their success in expressing opinions, persuading and engaging with readers, and the coherence of the text with respect to the task or context of the situation.

It was important to include the use of Appraisal resources as one of the criteria in the marking rubric because a number of studies have demonstrated the importance of Appraisal in quality essays. As we have seen, research on Appraisal (see, for

example, Derewianka 2007, Hood 2004, S. H. Lee 2006; 2007; 2013, Nakamura 2009) has revealed that quality academic essays are characterised by the use of a high number of Appraisal choices which are more ‘objective’; that is, entities are evaluated according to social value or worth (e.g., *their social networks are inadequate*) rather than more subjectively oriented direct expressions of feelings (e.g., *I hated it*) or judgement of people (e.g., *they are immoral*). Teachers’ understanding of the Appraisal system would help them recognise when students were expressing personal emotions rather than the more highly valued institutional framework of social feelings (Martin 2000). For this purpose, the intervention program provided some materials focusing on identification of Appraisal elements in academic writing found in high quality and low quality student essays.

**Table 4. 4 Interview excerpts on difficulties in marking vocabulary**

Researcher	Dina	Jelita	Ratna
<i>Now, of all those features of student essays, which one do you find most difficult to mark?</i>	Well, I think ... probably Vocabulary, because it really depends on student’s knowledge, I mean general knowledge when they use the Vocabulary.	Vocabulary. Some students tend to use very big words, without even checking whether it is suitable or not, but some other students they try to use simple vocabulary but their essays are really good.	I think ... specially for Vocabulary; I need to make sure whether it is the right diction, the right spelling or the right grammatical used. That’s the most difficult thing because probably because I’m not a native speaker of English,

When the participants in this study were interviewed on their perceived problems in marking student essays, they identified difficulties in marking around students’ vocabulary (see Table 4.4). As all of the participants were the teachers who had been teaching EAP classes at the institution for several years, the researcher was aware that the teachers had good field-related vocabulary related to the issues of globalisation. Thus, I drew their attention to separate the content words used by students (field-related vocabulary) from attitudinal words (tenor-related vocabulary),

which in this case belongs to the domain of the Appraisal system. By doing so, the teachers began to see students' attitudinal positions in their essays more clearly.

When designing a program, it is also important to consider the participants' demography and circumstances, and the present study is no exception. As all of the participants were adult learners, the researcher had to design a program suitable for adult learning. First of all, instructions should be customised as much as possible to the individual learner, with the trainer acting more as facilitator than teacher, drawing on teachers' experience wherever possible as a resource in the learning (Miles 2003). Using teachers' experience was made possible because of the pre-training data analysis of teachers' difficulties in marking student essays. Secondly, adult learning should have a real world focus and be rooted in what educators call "authentic" tasks whenever possible (Miles 2003). In relation to this study, the authenticity of the program was maintained through using real-life scenarios related to the participants' profession as EAP teachers and markers. Adult learning theory also stresses the importance of learner-relevance at every point in the learning process, the training program in this study endeavoured to connect to its participants' practices through using familiar materials such as students' sample essays, essays from the textbooks used in their institution and related to the genres they usually marked.

Another important factor to be considered before conducting the training program was the participants' availability. I recognised the participants' busy working hours. As a result, the training was conducted in short sessions of two hours over a period of four weeks, and was held on both weekdays and weekends, depending on the

participants' availability.

#### **4.3.3 Phase three – Implementing new practices**

This third phase of the study focused on the implementation of the SFL-informed rubric, particularly on its effects on inter-rater consistency level and the teachers' responses to the intervention. The phase involved sourcing data from teachers' marking practices when using the new rubric and from semi-structured interviews. This set of data was used to answer RQ3: *How do raters' marking practices shift after using SFL-informed criteria?* and RQ4: *How do raters respond to the SFL-informed criteria?* In order to see if there were any changes in teachers' marking and in the inter-teacher marking reliability scores when using the new rubric was measured using Stemler's (2004) consensus estimate counting formula. The result was then compared to their reliability when using the old rubric. One-to-one, semi-structured interviews were conducted for the purpose of gaining insights into teachers' opinions of the new rubric. As in Phase One, this interview technique was used because it enabled me to collect in-depth information by engaging with each teacher's perspective on the intervention (Patton 2002).

In Phase three the focus of the interviews was on the questions related to the SFL-informed marking criteria. In order to explore whether teachers' marking practices had changed as a result I asked them questions such as: "How do you find your rating when using the SFL marking criteria?". In order to explore the clarity of criteria in the new rubric I asked them such questions as: "Are the criteria provided easy to understand?", and in order to ascertain their perceptions of efficacy, I asked them questions such as: "Does this assessment help your marking of your student's

work more fairly?”. In order to find out their perceptions of the intervention, I asked them questions like: “What do you think about the rater-training program given?” (see Appendix C for interview guides). The main purpose of the second interviews was to gather information on teachers’ perceptions of the new marking rubric (the SFL-informed rubric) and the intervention program.

#### **4.3.4 Essay marking**

The samples of essay used for this study were obtained from Upper-Intermediate level EAP students who were studying at the language institution where this research took place. Students at this level can be expected to write short essays quite competently. The essays were part of their class assignment in the language school and were copied and retyped only after they had received scores from their teachers. To avoid identification of the students, the essays were retyped and given alphanumerical codes such as EAPs1, EAPs2 and so on. In retyping the student essays, I tried to retype them as closely as the original work; there was no correction made in terms of punctuation, spelling, vocabulary, and grammar mistakes. There were 20 essays written by 20 students on the same topic: *Globalisation can have negative impact on developing countries*.

The essay marking for this study was conducted in two phases: pre- and post-training. In the first phase the teacher participants were asked to mark one half of the 20 student essays using the marking rubric they normally used at the language school where this study took place. In Phase Three, they were asked to mark the other 10 essays using the new, SFL-informed rubric. The 20 essays were divided randomly into two bundles of ten essays in order to minimise the effects of teachers’

familiarity with the essays on scores they assigned to them. The participants were given choices about where and when they marked the essays as was normal practice, that is the essay marking activities were not bound by time and place. The purpose of this was to recognise the teachers' busy hours and to ensure that they were comfortable and familiar with the process and did not feel awkward with the tasks given.

#### **4.3.5 Ethical considerations**

There were several ethical considerations that this study had addressed before its commencement. Firstly, ethical approval from the University of Wollongong was obtained prior to beginning the data collection phase in Indonesia. The approval of the principal of the institution where the study was to take place was also gained. The second consideration was informed consent. The participants, the teachers who marked the essays and the students whose essays were used for this study, were firstly given consent forms before the commencement of data collection. The consent form clearly described the purposes, length, and benefits of the study to the participants in detail. The consent form also described the obligations of the teacher participants in the training program, and informed them of their rights to participate voluntarily (Creswell 2007, Keeves 1988).

Further, a relationship of reciprocity was formed in this study in that the teachers participated voluntarily and I was offering professional learning useful for individuals and the institution.

#### **4.3.6 Research relationships**

Educational research is often undertaken by people who have a history of involvement in the field of education in a capacity other than of researcher (Radnor 2002, p.30). This study was no exception. As I was ordinarily employed at the language school where this research was conducted, I, inevitably, had formed professional relationships with the participants prior to the research being conducted. This, to some extent, may have decreased the objectivity and generality of the findings (Lapan, Quartaroli et al. 2012). However, the intent of a qualitative inquiry is not to generalise to a population, but to develop an in-depth exploration of a central phenomenon (Creswell 2005, Hennink, Hutter et al. 2011, Merriam 1998). To best understand this phenomenon, the qualitative researchers purposefully or intentionally select individuals and sites (Creswell 2005, Hennink, Hutter et al. 2011). Likewise, as a qualitative study with a limited number of interviews and essays to be marked, this study did not try to generalise the findings to other educational and situational contexts in Indonesia. This study, instead, aimed to provide an in-depth analysis of participants' attitudes and agreement among them as EAP markers in a particular educational context.

#### **4.4 Triangulation of data**

In the attempt to obtain a broad understanding of the problems, and to construct credible interpretations and conclusions, the data was triangulated by using multiple data collection techniques such as interviews and marking practices (Merriam 1998, Stake 1995; 2010). The first technique was collecting data through interviews, which was conducted in a two-phase-research design: before and after the training

program. The purpose of the interviews was to yield direct quotations from participants about their experiences, opinions, feelings and knowledge (Patton 2002) with respect to marking practices, marking rubrics used, and the intervention.

Then, I did a ‘member-checking’ of the interviews to check the accuracy of the data, seek their approval, add additional information needed and clarify any ambiguities before being analysed (Kervin, Vialle, Herrington and Okely 2006, Merriam 1998, Stake 1995; 2010).

The second data collection technique involved collecting all the essays that had been marked by the teachers, both before and after the rater training was given. The first set of essays, for example, were analysed to identify the features of the essays that the raters least and most agreed upon and to measure the level of consensus estimate of inter-assessor reliability. Based on these analyses and the data gathered from the interviews conducted on Phase One, I then designed a new marking rubric, which was informed by SFL. Afterwards, the teachers participated in an intervention program, in which they were introduced to the model of the SFL approach to teaching language and its underpinning concepts, and the SFL-informed rubric. In the last stage, the participants were asked to mark the second group of ten essays by using the new rubric. Post-training data analyses examined the raters’ responses toward the redesigned rubric and their level of consistency in marking the second group of essays. The detailed analysis of the data can be seen below (see Section 4.5).

## **4.5 Data analysis tools**

The data comprised six hours of recorded interviews data (approximately 16,500 words) and 20 student essays (approximately 300 words each), 6,000 words in total. The analyses were recorded and stored as Microsoft Word and Microsoft Excel documents.

The analyses of the data were made up of two sets of analyses (linguistic and non-linguistic analyses). The non-linguistic analysis used in this study was the analysis of inter-rater reliability counting based on their consensus estimate. This consensus estimate is based upon the idea that raters should be able to come to reasonable agreement about how to apply the scoring rubric (Stemler 2004). This section describes each of the two groups of analyses.

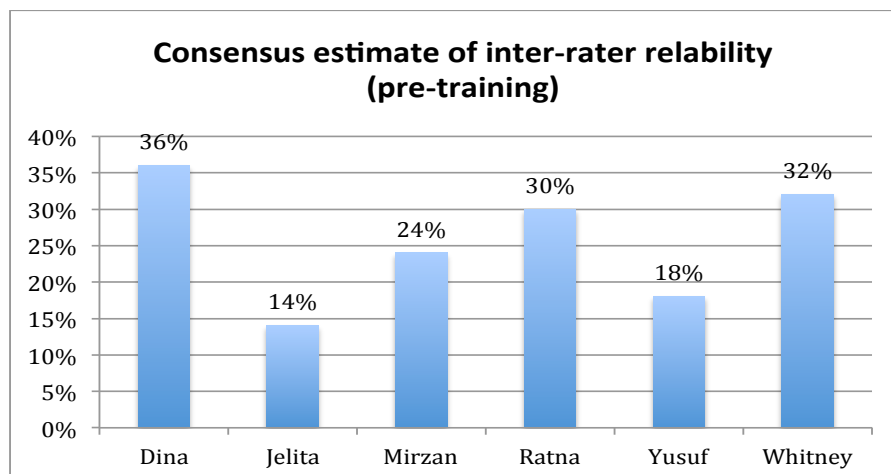
### **4.5.1 Consensus estimate of inter-rater reliability**

In order to be able to see whether the participants' marking practices have shifted after using SFL informed criteria (as in RQ3), the researcher needed to analyse teachers' level of reliability in marking student essays before and after the training was given. This would address the researcher's bias in interpreting the interview data, particularly those related to marking practices and marking consistencies. For this purpose, inter-teacher marking reliability needed to be counted.

As it has been discussed (see Chapter Two), inter-rater reliability can be defined as the extent to which an assessment would produce the same, or similar, score on two occasions or if given by two assessors (Martin and Rose 2007, Martin and White 2005). Following Stemler (H. D. Brown 2004, Gipps 1994, Stemler 2004, Weigle

2002), the percentage agreement figure was calculated by adding up the number of cases that received the same rating scores by both raters and dividing that number by the total number of essays rated by the two raters. This method of computation was advantageous because it was easy to calculate (could be done by hand) and is a commonly-used tool.

The range of scores assigned to each essay in this study was wide: the lowest possible score was zero (0) and the highest maximum score was twenty (20). As Stemler (2004) explains, the percent-agreement figure can be modified by broadening the definition of agreement through inclusion of the adjacent scoring categories on the rating scale. For example, on a rating scale with levels ranging from 1–7, assessors would not need to come to exact agreement about the ratings they assigned. So long as the ratings did not differ by more than one point above or below the other rater, then the two raters would be said to have reached consensus (Stemler 2004). Such an approach was applied in the consensus estimate counting in this study. The proportion or percentage of raters' consensus-reliability was then illustrated through visual representations, such as diagrams, to give clearer overall pictures of the trends emerging from the program (see Figure 4.2). These figures were then used to help the researcher interpret the qualitative data more thoroughly. These analyses could eventually help the researcher answer the overarching question of this study, which is: *To what extent does SFL-informed assessment criteria improve raters' consistency in an Indonesian EFL context?*



**Figure 4.2 Teachers' marking reliability when using the existing rubric**

### **4.5.2 Linguistic analysis**

Linguistic analysis was useful for a number of reasons. Firstly, it enabled me to look closely at the teachers' attitudes in their interviews because of the systematic interpersonal tools offered by SFL. Secondly, it allowed me to analyse closely the texts in order to design the new rubric by looking at the rhetoric patterns, expression of ideas, reader-writer interaction and the progression of information in student essays. Thirdly, it also allowed me to analyse closely the features of the essays to look at greater/lesser consistency before and after the intervention.

The linguistic analyses draw a range of language choices made at the levels of whole text, paragraph and sentence/clause. In addition, linguistic analysis of the students' texts was also conducted using lexical density and nominalisation counting of texts. As explained in the previous chapter, highly written texts tend to have higher lexical density and nominalisation than spoken texts (Eggins 2004, Halliday 1985b). As nominalisation deals with abstract knowledge, it is seen as a grammatical feature of the arguments of more mature writers (Knapp and Watkins

1994). Thus, analyses of student texts based on lexical density and nominalisation would reveal students' levels of maturity in writing. Included in the linguistic analyses were: genre analysis, Clause complex, lexical density and nominalisation, Theme analysis and Appraisal analysis. These analyses are described in the sections below.

#### **4.5.2.1 Genre analysis**

The purpose of genre analysis in this study was two-fold. In terms of student essays, it focused on the purpose of the essays (i.e., to argue for or against the impact of globalisation in developing countries) and stages used to achieve the purpose. In terms of teachers' marking, it helped inform the extent of changes in their marking practices. In order to see the changes, the essays that received the highest inter-assessor agreement, both when using the existing, non-SFL marking rubric and the SFL-informed rubric, were compared in terms of the scores given to each of the stage of argument genres. The stages used for Exposition were *Thesis ^ Arguments ^ Restatement* (Coffin 2004 p. 236, Martin 2009 p.13, Rose 2005, p. 9, Rose and Martin 2012 p. 272) and for Discussion *Issue ^ Sides ^ Resolution* (Knapp and Watkins 1994 p. 122, Rose 2005, p. 9, Rose and Martin 2012 p. 272). This analysis was used to answer RQ3. *How does raters' marking shift after using SFL-informed criteria?*

To provide a detailed picture of the raters' marking shift, Genre analysis was also conducted on students' argument essays. In this analysis, the essays that the teachers had marked were compared in terms of features that had the highest and lowest agreement scores. The genre analysis also enabled aspects of the consensus estimate

of reliability counts. The method of computing the agreement scores was through the use of the percentage of agreement figure, using Stemler's method of consensus estimate of reliability counts (2004). The detailed information of the consensus estimate counting was explicated above (see Section 4.5.1).

In order to identify the sources of assessors' inconsistency in marking, the features of the marking rubric used were analysed as they were named in the existing marking criteria. This was done to inform the researcher about the criteria which might need to be added, eliminated, altered or made clearer and easier for the assessors to understand. In addition to responding to answer RQ2, this analysis, along with the analysis of teachers' interviews, was used to inform a redesign of the rubric used for the post-training marking activities.

#### **4.5.2.2 Register analysis**

##### **The clause complex**

Functional theorists such as Eggins (2004), Halliday (1985b; 1994) and Martin (1985; 1991) have suggested that spoken language tends to be organised using more clauses chained together while in written language a large number of content words are typically packed into a single clause, and the grammatical relations between the clauses are usually very simple. Such analysis would inform the analyst/marker. It gives insight into how students are connecting ideas and controlling logical relation.

##### **Lexical density and nominalisation**

Lexical density counting was conducted to find out whether the high-scored, high inter-rater reliability text marked by the participants using the SFL-informed rubric

had more condensed information than the low-scored, high inter-rater reliability one. This counting of content word density was calculated by expressing the number of content carrying words (lexical items) in a text/sentence as a proportion of all the words in the text/sentence (Halliday 1985b). Content-carrying words include nouns, and the main parts of verbs, adverbs and adjectives. Non-content-carrying words include prepositions, conjunctions, auxiliary verbs and pronouns. To demonstrate the classification of the lexical and non-lexical items in a clause, consider for example: *Despite all the positive impacts mentioned above, globalisation has affected developing countries in negative ways*. This sentence consists of 10 lexical items (underlined) and 5 non-lexical items. For ease of comparison, the proportion of lexical density counting was multiplied by 100 in order to find the percentage. The formula is as follows:

$\text{Number of lexical items} : \text{Total number of words} \times 100 = \text{Lexical density}$
---

Nominalisation is a grammatical feature of mature writing (Knapp and Watkins 1994). The term refers to a process of turning elements that are not normally realized as nominal groups into nominal groups. Highly nominalised written text, then, allows a far greater proportion of the words in the text to be content carrying. Thus, written language generally has a much higher rate of lexical density than does spoken text (Christie and Derewianka 2008, Eggins 2004). However, as the essay's prompt was: *Globalisation can have negative impact on developing countries. Do you agree?* The nominal groups such as *globalisation* and *developing countries* were not included in lexical density and nominalisation counting.

## Theme analysis

In this study, Theme analysis of student essays was conducted to analyse the organisation of the message in student texts. Theme analysis was important for it could help reveal the coherent/incoherent organisation of ideas, and how they are unfolded in student essays. This was done through analysis of thematic progressions of ideas related to the field under discussion (the impacts of globalisation). Examples of Theme and Rheme are presented below:

**Table 4. 5 Theme-Rheme pattern**

<b>Theme</b>	<b>Rheme</b>
Globalisation	has given several advantages to developing countries (EAPs8a, 1).
The main advantage	is how every country is now connected to each other (EAPs8a, 2).
Globalisation	also causes westernisation (EAPs10a, 4).
Some of the people in developing countries	are not using local brands (EAPs10a, 5).
<b>Because of this</b>	many domestic industries are closed (EAPs10a, 7).

## Thematic progression

There are three types of Theme progressions, in which texts develop the ideas they present (Eggins 1994; 2004, Ravelli 2005). Following Eggins (2004, p. 324-326) and Ravelli (2005, p. 57), the thematic progression in students' texts can be classified into:

- (i) Theme reiteration: the text develops by repeating the same Theme and adding a new Rheme to it.
- (ii) The zigzag pattern: the text develops by picking up something from Rheme, and using it to start the following clause; an element which is introduced in the Rheme in clause 1 is promoted to become the Theme of clause 2.
- (iii) The multiple-Rheme pattern: the Theme of one clause introduces a number of different pieces of information, each of which is picked up and made a Theme in subsequent clauses.

An example of each thematic progression can be seen below:

## The Theme reiteration:

- CI 1. Industries and factories are another negative impact of globalisation.
- ↓
- CI 2. many industries and factories are built to fulfil human needs.
- ↓
- CI3. Many industries and factories mean an increase in the rate of pollution.

### The zig-zag pattern:

- CI 1. Another negative impact of globalisation is multinational power.
- CI 2. Multinational companies are spreading all over the world.

### The multiple-Rheme pattern:

- 
- ```
graph TD; CI1[CI 1. Globalisation has given several advantages to developing countries.] --> Center(( )); CI2[CI 2. The main advantage is how every country is now connected to each other.] --> Center; CI3[CI 3. Developing countries can gain access to the latest technology.] --> Center; CI4[CI 4. The other advantages are employment of human resources due to the arrival of multinational companies in developing countries.] --> Center;
```
- CI 1. Globalisation has given several advantages to developing countries.
- CI 2. The main advantage is how every country is now connected to each other.
- CI 3. Developing countries can gain access to the latest technology.
- CI 4. The other advantages are employment of human resources due to the arrival of multinational companies in developing countries.

## Bigger Waves: hyper and macroThemes

In addition to the Theme of a clause as discussed previously, Theme analysis can also be conducted at higher levels, such as the level of paragraph and of whole text, such as student essays, to reveal how layers of Theme development work together to create information flow, in which meanings are packed for ease of reading.

Martin and Rose (2007) propose the metaphorical term of ‘waves’ for this Theme development. The little wave of information at the clause level is called Theme, and

the bigger ones are called hyperTheme and macroTheme. While a Theme is a point of departure of a message telling what the clause is going to be about (Eggins 2004, Halliday 1994, Halliday and Matthiessen 2004), a hyperTheme predicts what each phase of discourse will be about. Traditionally, it can be seen as the ‘topic sentence’ of each body paragraph and it establishes expectations about how the text will unfold (Martin and Rose 2003; 2007). In the existing marking rubric, this falls under the criterion of Topic Sentence. The macroTheme, on the other hand, can be referred to the highest level of Theme predicting a set of hyperThemes or what the whole discourse is about (Martin and Rose 2003; 2007). This is the introductory paragraph of school rhetoric (Martin 1992, p. 437, Martin and Rose 2003; 2007). In the existing marking rubric, this is identified as General Statement and Thesis Statement.

Hierarchically, the simpler explanation of these layers of Themes would be: “The first paragraph orients the reader to the text. The topic sentences orient the reader to each paragraph. The topical Themes of clauses signal what the text is concerned with (the gist of the text)” (Butt, Fahey et al. 2000, p. 153). These layers of Themes are illustrated in Figure 4.3.

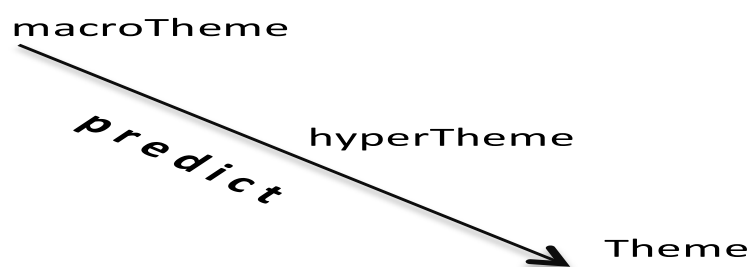


Figure 4.3 Layers of Themes (Martin and Rose 2007, p. 199).

## Appraisal analysis

In this study, the Appraisal system was used to analyse student essays and teachers' interviews. Analysis of student essays was conducted to inform how the Appraisal resources were employed in students' evaluations of the issue (the impacts of globalisation on developing countries), positions and resources used to engage with readers. As previously discussed, literature has shown the characteristic of quality academic essays in terms of Appraisal resources employed (see, for example, Derewianka 2007, Hood 2004, S. H. Lee 2006; 2013, Nakamura 2009). Thus, the intervention, which focused on promoting teachers' identification of Appraisal resources, would help enhance the quality of essay marking. In this study, two essays with the highest and lowest mean score were analysed and compared in terms of the Appraisal resources used in them. The analysis would contribute to the researcher's understanding of why some essays attracted higher marks.

In order to investigate the teachers' attitudes and values, an analytical tool that could uncover the linguistic form to reveal the different aspects of attitudes of the speakers was needed. Some studies such as Thomson et al (2006) and Tran (2011) have successfully used Appraisal as a tool of analysis of attitudinal aspect of spoken language such as in interviews. As explicated previously, an in-depth understanding

of raters' self-assessment, identified problems and perceptions can be best obtained through interviews. The transcripts of interviews used to answer RQ1 (*How do raters perceive the existing marking criteria?*) and RQ4 (*How do raters respond to the SFL-informed criteria?*) were analysed using Appraisal to 'get at' the teachers' attitudes and opinions regarding the existing and newly-developed marking criteria, and the features of student essays they find problematic to mark, as well as to inform the researcher about the rater training given.

Appraisal consists of three resources: ATTITUDE, ENGAGEMENT and GRADUATION. While ATTITUDE is concerned with feelings, opinions and judgments of behaviour, ENGAGEMENT is concerned with taking a stance and aligning with the reader and the discourse community. GRADUATION is concerned with grading phenomena. Using a model based on Coffin (2000, p. 178), a notational colour coding for Appraisal systems and sub-systems was adopted (see Appendix D). In addition to the colour coding system, the Appraisal systems/subsystems were also counted and presented on tables using abbreviations based on Martin and White (2005, p. 71), so as to give a more detailed picture of the evaluative words used (see Appendix D).

## **ATTITUDE**

In the present study, analysis of ATTITUDE resources was useful because it revealed feelings and attitudes in teachers' interviews and in student essays when discussing topics or fields. ATTITUDE consists of three semantic regions of feeling referred to as *Affect* (emotion), *Appreciation* (aesthetics) and *Judgment* (ethics). *Affect* encodes emotions and feelings, and its sources are conscious participants, including persons, human collectives, and institutions (Martin and Rose 2007, Martin and White 2005,

White 2012). This study adopted four sub-categories of Affect: *un/happiness*, *in/security*, *dis/inclination*, and *dis/satisfaction*. The *un/happiness* is concerned with the human's feeling of happy or sad, or feelings triggered by liking or disliking. The following comments came from one of the teachers about the new rubric, and revealed her positive opinions about the criteria used:

1. I really like the break down, ... (Whitney, 5)
2. I, especially also like the Communication Roles, the Tenor, (Whitney, 7).

*In/security* is related to emotions of peace and anxiety in relation to a human's environs. For example, the comments below made by two of the teachers when they were asked about the difficulties in using the existing rubric, reveal their concerns about it:

3. I'm not really clear about Readability (Dina, 10)
4. Sometimes I find it difficult to give the score (Yusuf, 3)

*Dis/inclination* is concerned with the expressions of emotive feelings related to 'fear' and 'desire' such as *afraid*, *anxious*, *worried*, *reluctant* and *unwilling*. *Dis/satisfaction* deals with feelings of achievement and frustration in pursuing goals. In Excerpts 5 and 6, for example, the teachers expressed their satisfaction with the new rubric, and at the same time making positive judgements of their marking capabilities:

5. I believe the level of consistency in marking my students' assignments has increased compared to before I got the training (Mirzan, 1).
6. It is easier for me in marking my students' essays rather than before (Ratna, 6).

*Appreciation* deals with meaning-construing evaluations of things, and it can be divided into three subsystems: *reaction*, *composition* and *valuation*. *Reaction* is

concerned with the impact of things that trigger emotion. The examples below were teachers' reactions to Appraisal and genre in the new marking rubric:

7. For the Appraisal word choices, I was introduced with a lot of new term and I have to remember them, so it's quite **difficult**. (Mirzan, 11).
8. what I find really **interesting** is that the last bit of genre (Dina, 21).

*Composition* is concerned with perceptions of order, complexity or balance. The first two excerpts below showed teachers' negative opinions about the construct of the old rubric. In contrast, the last excerpt showed a student's positive attitude towards the effect of globalisation.

9. ... the rubric **doesn't have a space** to assess the content (Dina, 37).
10. ... the first four criteria General Statement, Thesis Statement, Topic Sentence and Supporting Sentence should be only categorised into **one feature** (Yusuf, 19).
11. It opens a chance for **multinational** corporations to sell their products to international markets (EAPs1a, 5).

*Valuation* is related to cognition and it excludes reaction and composition. In the two examples below, the Thesis Statement and technology were valued positively by the teacher and the student respectively.

12. Probably the Thesis Statement because that's the **most important** part of an argument essay (Whitney, 20).
13. Developing countries can gain access to the **latest** technology (EAPs8a, 7).

*Judgment* is concerned with judging people's behaviour or character. *Judgment* can be divided into two groups: 'social esteem' and 'social sanction'. *Social esteem* has to do with how special, how capable or how dependable someone is. *Social sanction*, on the other hand, has to do with how truthful or how ethical someone is, and it leads to moral or legal praise and condemnation (Martin and Rose 2007, Martin and White 2005).

There are three sub-types of *social esteem*. *Normality* has to do with negative or positive evaluations of normal qualities of people. In Excerpt 14, one of the teachers made negative evaluation of her students, who usually forgot the importance of a Topic Sentence. In Excerpt 15, however, the teacher was making judgment of what she normally did when assessing her students' essays.

14. ... although the students **tend to forget** how important Topic Sentences are (Jelita, 4).
15. ... then I'll be **more forgiving** about the minor details (Whitney, 18).

*Capacity* has to do with positive or negative evaluations of people's capability or competence. As seen below, the teachers judge their marking capabilities when using the existing rubric differently; one appraised her capability negatively while another one appraised it positively.

16. ... it's **hard** for me to be **consistent** you know, (Dina, 9).
17. I am **consistent** with my marking, with the way I look at the essays, (Ratna, 8).

*Tenacity* is concerned with positive or negative evaluation of people's dependability in terms of determination. For example, Jelita in Excerpt 18 judged her students as lacking in willpower to study by calling them 'not-very studious'. Excerpt 19, in contrast, showed his determination in marking by saying 'try my best to be consistent'.

18. ... but some of the not-very **studious** ones they complain that, (Jelita, 24).
19. So I **try my best** to be **consistent** (Yusuf, 23).

*Social sanction* can be categorised into: *veracity* (how truthful or honest someone is) as in 20 and *propriety* (how ethical someone is) as in 21. For example, the comment in Excerpt 20 was made by one of the teachers when asked about the training program. The use of 'to be really honest' gave quality to her honesty to tell the truth.

In Excerpt 21, however, judgement was made in terms of ethics, which in this case was the teacher's abidance of the regulations of the school he worked for.

20. To be really **honest** to you, I didn't have time to explore what you have given (Dina, 42).
21. I try my best to **stick on** the rules ... the lists (Yusuf, 1).

### **Direct and indirect evaluations**

Evaluations in discourse can be represented in two options. One is 'direct inscription' through the use of attitudinal lexis, typically adjectives. The other one is through 'tokens' or ideational lexis that, indirectly, invokes our evaluation (Martin and White 2005).

22. ... so we have to use our own discretion to check where we have to put grammar (Jelita, 8).
23. Just like I said to you before, because from the previous rubric I don't have ... the rubric doesn't have a space to assess the content (Dina, 32).

In Excerpts 23 and 24, for example, there is no attitudinal lexis used to evaluate the existing, non-SFL marking rubric. However, the teachers' negative evaluations can be deduced from the ideational meanings such as 'so we have to use our own discretion' and 'does not have space to assess the content'.

### **ENGAGEMENT**

The second appraisal subsystem used in analysing students' and teachers' texts is ENGAGEMENT, which enabled me to consider the strengths of teachers' opinions. Analysis of ENGAGEMENT in teachers' talk would give insight into the strength of their feelings about marking, the rubric and their students, while in the students' texts it is an indicator of how a mature argument is prosecuted. ENGAGEMENT

provides two broad categories of presenting propositions. One is *monogloss* – often referred to as ‘bare assertion’ – by which the writer/speaker makes no reference to or allows no space for other voices and viewpoints in their texts (White 2006). The statement in Excerpt 24 was made by the teacher when asked about the criteria in the SFL-informed rubric. In Excerpt 25, a similar bare assertion was found in a student’s essay when evaluating the effect of globalisation on third-world countries.

24. The features **are** easy to understand (Mirzan, 14).

25. Wages **are** reduced and safety standards **are** ignored (EAPs8a, 12).

The other category is *heterogloss*, by which the writer/speaker makes some reference to other voices and viewpoints and allows for alternative positions (Martin and White 2005, White 2001). In the present study, some of the heteroglossic options indicating a space for dialogic alternatives were identified as: *deny*, *counter*, *concur*, *pronounce*, *endorse* and *entertain*.

*Deny* is a heteroglossic resource introducing the alternative position by rejecting it. In the example below (see Excerpt 26), one of the teachers made a negative comment about the old rubric by denying its focus on the content of the essay. Similarly, in Excerpt 27 one of the students denied the benefit of globalisation in her essay.

26. However, in the old version, I mean the existing one, the content is **not** the focus (Jelita, 15).

27. The benefit of globalization is **not** universal (EAPs8a, 22).

*Counter* is a heteroglossic resource, which replaces or supplants a proposition that would have been expected in its place. The comment below was made by one of the

teachers regarding her leniency in essay marking, using an element of *counter*, ‘but’ to defend her position as being ‘less forgiving’.

28. **But** if they show that they’ve just done it the night before, they haven’t given any thought and it’s a mess, that’s when I’m less forgiving (Whitney, 19).

*Concur* is a heteroglossic resource which overtly expects the putative reader/speaker to have the same position or knowledge as the writer/speaker. The comment below was made by a teacher, Yusuf, when asked about the usefulness of the intervention, which shows his assumption that others would have opinions similar to his.

29. After attending the training, **of course** I find it quite helpful (Yusuf, 1).

*Pronounce* is a heteroglossic resource involving explicit authorial emphases or interventions. The two comments below explicitly show the teachers’ positions regarding the helpfulness of the SFL-informed rubric to enhance consistency and to mark the content of students’ essays.

30. Well, **it certainly helps me** to be more consistent (Dina, 1).

31. **Sure! Yeah, absolutely!** (Ratna, 23).

*Endorse* is a heteroglossic resource presenting an authorial position aligning with external sources. In the excerpt below the student grounded in his/her position as the writer of the essay by sourcing from some samples he/she called ‘condition’.

32. This condition clearly **shows** that there is a competition between international company and local company in developing countries (EAPs3a, 13).

*Entertain* is a heteroglossic resource presenting the authorial voice/position as one of a number of possible positions. The first comment below was made by a teacher regarding the intervention, which she thought was slightly too short. The second

comment was written by a student regarding one of the negative impacts of globalisation.

33. Well, *maybe* the time, *perhaps*, more extensive (Ratna, 17).
34. Polluted gas which is freely disposed in the air *might* cause serious damage to citizens' health (EAPs8a, 17).

## GRADUATION

The third resource of Appraisal used in the study was GRADUATION. It was important in the study because it can inform the strengths of students' and teachers' evaluations in terms of word choices made, whether the intensity of the words used were low, neutral or high. GRADUATION is concerned with grading phenomena, and it consists of two sub-systems of grading (*force* and *focus*). Force has to do with the assessment of amount and intensity of an evaluation by amplifying (raising) or down-toning (lowering). The scale that deals with amount is called quantification. There are two prominent sub-types under quantification: quantity (number or mass) and extent (scope in time or space). Intensification, on the other hand, deals with intensity of qualities and processes. Intensification can be further divided into isolating (i.e., the use of adverbs such as 'quite' or 'very') as in Excerpt 33, and infusing (i.e., the use of gradable lexical items such as 'scared > terrified > petrified') such as in Excerpt 34.

35. And then, what I find *really interesting* is that the last bit of Genre (Dina, 20).
36. *Definitely* the timing (Dina, 39).
37. The rich are getting *richer* (EAPs8a, 29).
38. In summary, globalisation in developing countries results in *stronger* international power (EAPs3a, 22).

The second subsystem, *focus*, is concerned with non-scalable categories of values. Instead of scaling or grading (like *force*), *focus* functions like the lens of a camera, to sharpen or to soften boundaries.

- 39. They're **kind of** memorizing by rote, (Whitney, 29).
- 40. They're **sort of** like in the middle not like straight to the point (Dina, 14).
- 41. The **main** impact is how every country now connected each other (EAPs8a, 6).

## 4.6 Summary

This chapter has described the research design, the research setting, the participants and the data. It also has detailed the design and the implementation of the intervention program. It has pinpointed the explicitness of the new marking rubric used in marking student essays. This chapter also has elaborated aspects of linguistic and non-linguistic analytical tools used around data related to both teachers' perceptions, attitudes and marking practices, and to students' argument essays. The next chapter outlines the results of the analysis of a successful text, how it was used to inform the development of the new rubric and how it was implemented in the intervention.

## **Chapter Five**

### **Designing SFL-informed rubrics**

The purpose of this chapter is to present the analysis of what was valued in students' texts, which was drawn on to inform the design of the new marking rubric and the intervention. The previous chapter described the research method used to investigate teachers' assessment practices and views. It outlined the significant contribution of SFL theory (Halliday 1978; 1985a, Halliday and Matthiessen 2004, Martin 1984; 2005) and its systematic tools for the shaping of the intervention and the analyses of data. It described the three phases of data collection: Phase One focused on mapping the existing assessment practices; Phase Two concerned the redesigning of a new assessment tool and its application through an intervention; and Phase Three involved implementing and reflecting on post-training marking practices. This chapter, Phase Two, describes the shared features of student texts that can be used to inform the design of the new rubric and intervention.

#### **5.1 Analysis of two sample essays**

SFL theory posits that language use can be identified by its social purpose or genre (see, for example, Martin 1992; 2005, Martin and Rose 2008, Martin and Rothery 1993) and based on the situation in which it occurs or is registered (see, for example, Halliday 1975, Halliday 1978, Halliday and Hasan 1985, Martin 1984) (see Chapter Three, Section 3.3). Informed by SFL theory, one successful student's essay and a sample successful essay from an IELTS (Jakeman and McDowell 2003) preparation book were analysed in terms of genre and register variables in order to identify

features inherent in texts judged as effective by markers. These features then informed the design of the new rubric and intervention. The IELTS sample was selected because one of the purposes of the EAP course in the language school where the study was conducted was to prepare students either to join an IELTS preparation course or take an IELTS exam afterwards. Thus, the essays selected for the study were, for example, similar to the writing tasks given in the IELTS preparation classes in terms of the text types used (exposition and discussion essays) (Moore and Morton 1999) and the size of the essay (approximately 250-300 words). Both EAP and IELTS courses were similar in terms of the intended students or test-takers in that both programs were designed for non-native English-speaking students who plan to study in English-medium countries at university level (Hamp-Lyons 2001).

The first step of this analysis was to identify the staging and phases of each essay. This analysis adopted the conventions of stages and phases after SFL-genre theorists such as Coffin (2004), Martin (2009), Rose (2005) and Rose and Martin (2012).

### **5.1.1 Genre analysis**

As discussed in Chapter Three, the study adopts the SFL approach to genre, which positions genre at the broader context (context of culture), and is seen as a realisation of the social purpose of language use (see Section 3.2.1). For example, the purpose of essay writing in the language course where the study took place was to put forward a point of view (Argument) or to argue a case for two or more points of views (Discussion). The essay below, which is the IELTS example, responded to the following prompt: *Should parents be obliged to immunise their children against*

*common childhood disease? Or do individuals have the right to choose not to immunise their children?*

### A highly-rated IELTS essay

| Stage and Phase                     | Text                                                                                                                                                                                                                                                                                                                                                            |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis<br>(information)<br>position | <sup>2</sup> Some people argue that the state does not have the right to make parents immunise their children. However, I feel the question is not whether they should immunise but whether, as members of society, they have the right not to.                                                                                                                 |
| Argument 1<br>topic                 | Preventative medicine has proved to be the most effective way of reducing the incidence of fatal childhood diseases.                                                                                                                                                                                                                                            |
| elaboration                         | As a result of the widespread practice of immunising young children in our society, many lives have been saved and the diseases have been reduced to almost zero.                                                                                                                                                                                               |
| Argument 2<br>topic                 | In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had immunity, the diseases spread easily.                                                                                                                                                                                                 |
| elaboration                         | Diseases such as dysentery were the result of poor hygiene but these have long been eradicated since the arrival of good sanitation and clean water. Nobody would suggest that we should reverse this good practice now because dysentery has been wiped out.                                                                                                   |
| Argument 3<br>topic                 | Serious diseases such as polio and smallpox have also been eradicated through national immunisation programmes.                                                                                                                                                                                                                                                 |
| elaboration                         | In consequence, children not immunised are far less at risk in this disease-free society than they otherwise be. Parents choosing not to immunise are relying on the fact that the diseases have already been eradicated. If the number of parents choosing not to immunise increased, there would be a similar increase in the risk of the diseases returning. |
| Restatement<br>review               | Immunisation is not an issue like seatbelts which affect only the individual. A decision not to immunise will have widespread repercussions for the whole of society and for this reason, I do not believe that individuals have the right to stand aside.                                                                                                      |
| restate                             | In my opinion immunisation should be obligatory.                                                                                                                                                                                                                                                                                                                |

---

<sup>2</sup> Jakeman, V. and McDowell, C. (2003). *Insight into IELTS extra: the Cambridge IELTS course*. Cambridge: Cambridge University Press.

The essay below, which was rated by the teachers, responded to the following prompt: *Globalisation can have negative impact on developing countries. Do you agree?*

#### Essay highly rated by teachers

| Stage and Phase                  | Text                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis<br>( <i>information</i> ) | In the past 20 years, globalisation has brought people closer. It is the process of opening up domestic economies to foreign competition and investment (EAP-4, p.64).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <i>position</i>                  | Globalisation brings several negative impacts to the economies of developing countries.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Argument 1<br><i>topic</i>       | One of the negative impacts of globalisation is tax evasion.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <i>elaboration</i>               | Tax avoidance by some multinational companies costs the developing nations approximately \$160 billion every year (Peake, 2011). Moreover, the profits of the developing countries have extremely decreased as a result of the tax dodging by some multinational companies.                                                                                                                                                                                                                                                                                                                                                                            |
| Argument 2<br><i>topic</i>       | Secondly, globalisation leads to exploitation of human resources.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| <i>elaboration</i>               | Most of the multinational companies set up their factories in developing countries such as Indonesia, China and India in order to produce cheap goods while hiring cheap labour. The employees in some multinational companies have to work overtime but receive low wages. Furthermore, the celebrities who promote the goods are paid more than the workers who have worked very hard to produce the goods. For example, the celebrities get paid about \$1 million for advertising a product whereas the workers get only \$5 per day to produce it. Therefore, the number of poor people is increasing due to the exploitation of human resources. |
| Restatement<br><i>review</i>     | In conclusion, the negative impacts of globalisation have affected the economies in developing countries, and they were very intense.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <i>restate</i>                   | Two of the most serious impacts are tax evasion and exploitation of human resources. Tax evasion by the multinational companies and exploitation of human resources have caused poverty in developing countries.                                                                                                                                                                                                                                                                                                                                                                                                                                       |

The essays above are instances of an argument genre, in particular an analytical exposition (Martin 1985), the purpose of which is to argue for a particular point of

view of an issue (Macken-Horarik 2002, Martin 1985) and to persuade the reader(s) to think in particular ways (Coffin 2004, Unsworth 2000). In this case, the writer of the IELTS-rated essay was trying to persuade the reader (that is, the examiner who marked the essay) that immunisation should be made obligatory. Similarly, the writer of the second essay was trying to convince the readers (the teachers who marked the essay) that globalisation had negative effects on third world countries. The schematic structure of the two texts is similar, characterising common elements of the argument genre, as follows:

- **Thesis.** The writers of the two essays above introduce the issues and their respective position or point of view. In the introductory paragraph of the IELTS-rated essay, for example, the writer states his/her position using a conjunction, *however*, as in: *Some people argue that the state does not have the right to make parents immunise their children. However, I feel the question is ...* This functions to counter other people's opinions that government has no right to make parents immunize their children. Similarly, in the teacher-rated essay, the writer's position is explicitly stated: *Globalisation brings several negative impacts to the economies of developing countries.* Based on the statement given in the Thesis, it is anticipated that the arguments will be built on the negative impacts of globalisation on the economies of third world countries.
- **Arguments.** Several arguments are used to support the Thesis in both texts. In the first essay, three arguments for immunisation are stated in the topic sentences followed by some elaboration to support the Thesis made. These arguments are: (1) *Preventative medicine has proven to be the most effective way of reducing the incidence of fatal childhood diseases;* (2) *In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had immunity, the diseases spread easily;* and (3) *Serious diseases such as polio and smallpox have also been eradicated through national immunisation programmes.* In the second essay, two topic sentences show concern for the negative effects of globalisation stated in the Thesis stage. These are: (1) *One of the negative impacts of globalisation is tax evasion;* (2) *Secondly, globalisation leads to exploitation of human resources.*

- **Restatement.** The writers review the issues raised and restate the particular point of view adopted (Rose and Martin 2012). In the IELTS-rated essay, the topic sentence reviews the points previously made in the Arguments stage. The last clause, *In my opinion immunisation should be obligatory*, functions to reiterate the writer's position on the issue. Similarly, in the teacher-rated essay, the review phase, *In conclusion, the negative impacts of globalisation have affected the economies in developing countries, and they were very intense*, functions to distil the points that the writer has made (Rose and Martin 2012), and at the same time draws conclusion from what has been argued in the previous stages. The restate functions to reiterate the writer's position in that globalization negatively affects the economies of third-world countries.

Based on the assessment scores given, it can be said that the two essays were successful in achieving their purposes. The first essay received a band score of 9, which is the highest possible score assigned to an essay on the IELTS test. The second essay, from the corpus collected for the study, received a mean score of 16.1 out of a maximum of 20. This essay received the highest score among the other essays assessed and the analysis suggested it was highly successful. Although most of the teachers regarded this essay as a good and successful argument essay, they tended to disagree with each other on the exact score it should have received. Indeed, the essay was one of the essays on which the teachers had the least consistency. This is indicated by the fairly low inter-rater reliability counts of the essay, which received only six similar scores out of a maximum of 30 recurrences. In terms of the existing marking rubric used, the variability of scores can be seen as an indication of the lack of explicitness of the criteria for success and what these looked like in the student texts. In order to make criteria more visible or explicit, identifying the valued features of student essays is an important contribution of the research described in this thesis. Such an analysis, as that described here, should

address the three metafunctions of language discussed in the previous chapter (see Chapter Three, Section 3.3.2) in order to reveal how successful texts are structured to unfold experience, reveal patterns of interactions between the writer(s) and the reader(s) and uncover the organisation or progression of information (how the texts hang together) to create cohesive and coherent essays (Eggins 1994; 2004).

### 5.1.2 Analysis of the ideational function of text: the clause complex system

The clause complex allows the analyst to account for the functional organisation of sentences (Halliday 1985a; 1994). As it consists of information about people, processes and things, the clause complex system is regarded as a content function of language, and thus belongs to the field of the text. Similarly, in the study, clause complex analysis was conducted to see how each sentence was constructed and to look closely at how clauses or ideas were combined to make various kinds of meaning. This would tell students' capabilities to pack information in noun groups/phrases and how they were systematically connected by processes (verbs) to reveal an experience or a happening. For this purpose, sentences were divided into clause simplexes (Simplex) and clause complexes (Cl. C). The notation conventions for the analysis are indicated below<sup>3</sup>.

In the IELTS-rated essay, clause simplexes and clause complexes are strategically employed to build arguments. Of the 14 clauses identified (see Table 5.1), the majority (approximately 65%) are organised as clause complexes. At first glance,

---

<sup>3</sup> The notation convention: double chevrons are for interrupting clause, double square brackets for embedded clause, and brackets with a caret for ellipsis. The dependent clauses are in **bold** type. Altogether, the coding system for this part is as follows: << interrupting clause>>, [[embedded clause]], (^ ellipsis) and **dependent clause**.

this high number of clause complexes seems to indicate that the essay is more influenced by speaking. As explicated earlier (see Chapter Four, Section 4.7.1.1) some functional theorists such as Eggins (1994; 2004), Halliday (1985b; 1994) and Martin (1985; 1991) have demonstrated that spoken language tends to be intricately organised around the clause in the form of clause-chains to accommodate its lexical items. In written language, on the other hand, a large number of content words are packed into a single clause in a generally simple grammatical relation. However, it can be argued that the writer used simplexes sparingly for particular purposes.

**Table 5.1 Clause combination in the IELTS-rated essay**

| Stages      | Simplexes | Complexes |
|-------------|-----------|-----------|
| Thesis      | -         | 2         |
| Argument 1  | 1         | 1         |
| Argument 2  | -         | 3         |
| Argument 3  | 2         | 2         |
| Restatement | 2         | 1         |
| Total       | 5         | 9         |

In Argument stages, for example, clause simplexes occur mainly in the *topic* (*hyperTheme*) to state the claim authoritatively:

*Preventative medicine has proven to be the most effective way of reducing the incidence of fatal childhood diseases.*

*Serious diseases such as polio and smallpox have also been eradicated through national immunisation programmes.*

Similarly, in the Restatement stage, clause simplexes occur in the *review* and *restate* phases to reiterate concisely the themes of the argument and reaffirm the writer's position:

*Immunisation is not an issue like seatbelts [[which affect only the individual]].*

*In my opinion immunisation should be obligatory.*

The use of clause complexes, on the other hand, can be seen as indicative of the writer's capability to deliberately unpack what might otherwise be too concisely and metaphorically packaged information in the text (Ventola 1996). In other words, the writer's unpacking of information can be seen as an attempt to facilitate the reading process (Hinds 1987).

### Highly-rated IELTS essay

| Stage and Phase  | Clause combination |       | Content                                                                                                              |
|------------------|--------------------|-------|----------------------------------------------------------------------------------------------------------------------|
| Thesis position  | Cl. C I            | cl. 1 | Some people argue                                                                                                    |
|                  |                    | cl. 2 | <b>that the state does not have the right to make parents immunise their children.</b>                               |
|                  | Cl. C II           | cl. 3 | However, I feel the question is not [[whether they should immunize]]                                                 |
|                  |                    | cl. 4 | but whether, as members of society, [[they have the right not to]].                                                  |
| Argument 1 topic | Simplex I          | cl. 1 | Preventative medicine has proven to be the most effective way of reducing the incidence of fatal childhood diseases. |
| elaboration      | Cl. C I            | cl. 2 | As a result of the widespread practice of immunising young children in our society, many lives have been saved       |
|                  |                    | cl. 3 | and the diseases have been reduced to almost zero.                                                                   |
| Argument 2 topic | Cl. C I            | cl. 1 | In previous centuries children died from ordinary illnesses such as influenza and tuberculosis                       |
|                  |                    | cl. 2 | and because few people had immunity,                                                                                 |
|                  |                    | cl. 3 | the diseases spread easily.                                                                                          |
| elaboration      | Cl. C II           | cl. 4 | Diseases such as dysentery were the result of poor hygiene                                                           |
|                  |                    | cl. 5 | but these have long been eradicated since the arrival of good sanitation and clean water.                            |
|                  | Cl. C III          | cl. 6 | Nobody would suggest                                                                                                 |
|                  |                    | cl. 7 | <b>that we should reverse this good practice now [[because dysentery has been wiped out]].</b>                       |
| Argument 3 topic | Simplex I          | cl. 1 | Serious diseases such as polio and smallpox have also been eradicated through national immunisation programmes.      |
| elaboration      | Simplex II         | cl. 2 | In consequence, children [[not immunized]] are far less at risk in this disease-free society than                    |

|                              |            |       |                                                                                        |
|------------------------------|------------|-------|----------------------------------------------------------------------------------------|
|                              |            |       | they otherwise be.                                                                     |
|                              | Cl. C I    | cl. 3 | Parents [[choosing not to immunize]] are relying on the fact                           |
|                              |            | cl. 4 | <b>that the diseases have already been eradicated.</b>                                 |
|                              | Cl. C II   | cl. 5 | <b>If the number of parents choosing not to immunise increased,</b>                    |
|                              |            | cl. 6 | there would be a similar increase in the risk of the diseases returning.               |
|                              |            |       |                                                                                        |
| Restatement<br><i>review</i> | Simplex I  | cl. 1 | Immunisation is not an issue like seatbelts [[which affect only the individual]].      |
|                              | Cl. C I    | cl. 2 | A decision not to immunise will have widespread repercussions for the whole of society |
|                              |            | cl. 3 | and for this reason, I do not believe                                                  |
|                              |            | cl. 4 | <b>that individuals have the right to stand aside.</b>                                 |
| <i>restate</i>               | Simplex II | cl. 5 | In my opinion immunisation should be obligatory.                                       |

In contrast to the essay above, in the teacher-rated essay, many more clause simplexes than complexes are employed to effectively build argument. As seen in Table 5.2 below, the majority (78%) of the clauses are simplexes while the rest 22% are in complex clauses.

**Table 5.2 Clause combination in the teacher-rated essay**

| Stages      | Simplexes | Complexes |
|-------------|-----------|-----------|
| Thesis      | 3         | -         |
| Argument 1  | 3         | -         |
| Argument 2  | 3         | 2         |
| Restatement | 2         | 1         |
| Total       | 11        | 3         |

The prevalence of simplexes in the teacher-rated essay is typical of written academic text. The text is characterised by a number of single clauses consisting of one or two nominal groups with simple grammatical relations between them (Halliday 2004, Martin 1985). The use of simplexes in the text functions to create authoritative voice. In the Thesis stage, for example, the simplex formulation realises the writer's assertiveness in the *position* he/she opts for, and at the same time, defines the topic

to be discussed (i.e., the negative impacts of globalisation on the economies of third-world countries):

*Globalisation brings several negative impacts to the economies of developing countries.*

Similarly, in the Argument stages, the simplexes employed in the *topic* (*hyperTheme*) shapes the writer's authoritative voice in that the points of argument are conveyed confidently and in a non-modalised form:

*One of the negative impacts of globalisation is tax evasion.*

*Secondly, globalisation leads to exploitation of human resources.*

Such sentences are simple grammatically but not semantically. The nominal groups are complex bundles of packages of information. As we shall see, this is a grammatical feature of written language. The writer's use of clause complexes, on the other hand, can be seen as his/her cautious way of unpacking the information. For example, in the second Argument stage of the essay, *Secondly, globalisation leads to exploitation of human resources*, the condensed information created through grammatical metaphor and nominalisation is unpacked to generate a less abstract and clearer message for the reader (Hinds 1987). The unpacking of the metaphorical noun (*exploitation of human resources*) can be seen in the *elaboration* phase of the essay, below:

*The employees in some multinational companies have to work overtime but (^they) receive low wages.*

*For example, the celebrities get paid about \$1 million for advertising a product whereas the workers get only \$5 per day to produce it.*

### Essay highly rated by teacher

| Stage and Phase                 | Clause Combination |       | Content                                                                                                                                                     |
|---------------------------------|--------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis<br>( <i>background</i> ) | Simplex I          | cl. 1 | In the past 20 years, globalisation has brought people closer.                                                                                              |
|                                 | Simplex II         | cl. 2 | It is the process of opening up domestic economies to foreign competition and investment (EAP-4, p. 64).                                                    |
| <i>position</i>                 | Simplex III        | cl. 3 | Globalisation brings several negative impacts to the economies of developing countries.                                                                     |
| Argument 1<br><i>topic</i>      | Simplex I          | cl. 1 | One of the negative impacts of globalisation is tax evasion.                                                                                                |
| <i>elaboration</i>              | Simplex II         | cl. 2 | Tax avoidance by some multinational companies costs the developing nations approximately \$160 billion every year (Peake, 2011).                            |
|                                 | Simplex III        | cl. 3 | Moreover, the profits of the developing countries have extremely decreased as a result of the tax dodging by some multinational companies.                  |
| Argument 2<br><i>topic</i>      | Simplex I          | cl. 1 | Secondly, globalisation leads to exploitation of human resources.                                                                                           |
| <i>elaboration</i>              | Simplex II         | cl. 2 | Labourers [[[working in multinational companies [[which are set up in developing countries such as Indonesia, China and India]] are generally underpaid]]]. |
|                                 | Cl. C. I           | cl. 3 | The employees in some multinational companies have to work overtime                                                                                         |
|                                 |                    | cl. 4 | but (^they) receive low wages.                                                                                                                              |
|                                 | Simplex III        | cl. 5 | Furthermore, the celebrities [[who promote the goods]] are paid more than the workers [[[who have worked very hard]] [[to produce the goods]]].             |
|                                 | Cl. C. II          | cl. 6 | For example, the celebrities get paid about \$1 million for advertising a product                                                                           |
|                                 |                    | cl. 7 | whereas the workers get only \$5 per day to produce it.                                                                                                     |
| Restatement<br><i>review</i>    | Cl. C I            | cl. 1 | In conclusion, the negative impacts of globalisation have affected the economies in developing countries,                                                   |
|                                 |                    | cl. 2 | and they are very intense.                                                                                                                                  |
| <i>restate</i>                  | Simplex I          | cl. 3 | Two of the most serious impacts are tax evasion and exploitation of human resources.                                                                        |
|                                 | Simplex II         | cl. 4 | Tax evasion by the multinational companies and exploitation of human resources have caused poverty in developing countries.                                 |

As seen above, clause combinations are effectively employed in the two essays for both building argument and rhetorical organisation. The most common selection typical of written academic text is clause simplex (Halliday 1998, Martin 1991). In each clause simplex identified above, two nominal groups are generally related as a single relational clause such as *is*, *are* and *be*. In other words, each of the simplexes typically consists of one nominal group plus one verbal group plus a second nominal group, and forms very simple grammatical relations between the groups (Martin 1985; 1991, Ravelli 1988). Although simple in structure, as Halliday (1998, p. 207) points out, such clause combination is ‘the favourite clause type of English scientific writing’, and such clauses “are the most critical in the semantic load that they carry in developing scientific argument”.

Each sequence of the two nominal groups related as a single relational clause has a significant role. In the Thesis stage, for example, a series of clause simplexes functions to help define the topic and delineate its scope. The relationship between them, construed as relational and existential processes, is in bold face:

*It **is** the process of opening up domestic economies to foreign competition and investment.*

In the Argument stages, the single clauses function to carry and develop arguments as well as their interpretations.

(in Argument 1)

*One of the negative impacts of globalisation **is** tax evasion.*

(in Argument 2)

*Globalisation **leads to** exploitation of human resources.*

(in Argument 3)

*Children not immunised **are** far less at risk in this disease-free society than they otherwise be.*

In the Restatement stage, single clauses appear in the review and restate phases to allow concise reiteration of the arguments stated in the body paragraphs of the essay:

*Two of the most serious impacts **are** tax evasion and exploitation of human resources.*

*Immunisation **is not** an issue like seatbelts which affect only the individual.*

*In my opinion, immunisation **should be** obligatory.*

The simplexes above are rich with metaphors of relationship of causality and conditionality. If the metaphors were unpacked for a more congruent version, each clause simplex would become a clause complex consisting at least two clauses in hypotactic interdependency: *Globalisation has serious impacts on developing countries because it makes it easy to evade taxes and exploit human resources; Because multinational companies evade taxes and exploit human resources, people in developing countries become poor; If you are not immunised, you will affect other people when you are sick; and Because of that I think everybody should be immunised.* Thus, it is sensible to say that the two essays demonstrate control of mature writing. Furthermore, when ideas are packed into this simplex structure there may usually be a very high density of lexical words (content-carrying words such as nouns, the main parts of verb, adverbs and adjectives). As Halliday (1998) argues, the average lexical density for spontaneous-spoken English barely exceeds two

lexical words per clause, it can be expected that written texts are more lexically dense than the spoken ones. This is because in written texts, the information load of message is carried out by participants (Ravelli 1988) in the form of grammatical metaphor i.e., when experience is construed incongruently in the form of nominal groups (Halliday 1985a, Halliday and Matthiessen 1999, Ventola 1996):

**Table 5.3 Nominal groups in grammatical metaphor**

| <b>nominal group (incongruent/metaphorical)</b>                                                                                                              | <b>clause (congruent/literal)</b>                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> <li>•the terrible exploitation of human resources</li> <li>•the devious tax evasion by multinational companies</li> </ul> | <ul style="list-style-type: none"> <li>•People work to produce goods and services are exploited terribly.</li> <li>•Companies which operate around the world deviously avoid paying tax.</li> </ul> |

Grammatical metaphors enhance the topic or field of the discourse. As seen in the two metaphorical examples above, experience is not realised as a clause or series of clauses but as a nominal group compacted with extra information. As a result, the topic or field of the discourse becomes more focused as the participants, realised by nominal groups in the clause, have been richly packed with information realised by pre- and/or post-modifiers (Ventola 1996). The nominalisation of processes such as *exploit* – *exploitation* and *evade* – *evasion* is a major factor in being able to manage reasoning required in the writing of an academic essay (Christie and Derewianka 2008).

In the present study, as in Christie and Derewianka (2008), the terms *nominalisation* and *grammatical metaphor* are used interchangeably with a distinction to be made. Nominalisation is a broad overarching term that can be used to refer to the phenomenon of construing various meanings in nominal form and may involve grammatical metaphor. The probe for grammatical metaphor is to ask whether the

nominalised word can be unpacked to “a plausible, congruent alternative” (Christie and Derewianka 2008, p. 26). In the two examples shown in Table 5.3 above, for example, the nominalisation involves grammatical metaphor (*the terrible exploitation of human resources* and *the devious tax evasion by multinational companies*) and can be unpacked into more congruent/literal forms (e.g., *Human resources are exploited terribly* and *Multinational companies deviously avoid paying tax*). In the present study, the grammatical metaphor analysis would inform the rubric design in terms of the field (content of the essay) as well as the teachers’ identification of this feature in the intervention.

### 5.1.3 Lexical density and nominalisation

The literature (see, for example, Halliday 1985b, Halliday and Hasan 1985, Martin 1985, Ravelli 1988) has shown that written texts are more lexically dense than the spoken ones. Even within numerous written texts, some are more lexically dense than others. Ure’s study in 1971 (cited in Ventola 1996), showed that spoken texts had a lexical density of fewer than 40%, a text from school textbook 50.2% and an academic article 52.8%. In this analysis, lexical density is underlined and nominalisation is boxed.

Based on the lexical density counts, each of the two essays (The IELTS-rated and teacher-rated essays) has quite a high level of density with an average of 51.2% and 59% respectively. Lexical density of the IELTS-rated essay is slightly below that of the academic article. In contrast, the lexical density of the teacher-marked essay is over.

**Table 5.4 Lexical density and nominalisation in the two essays**

| Genre stages | Lexical density |               | Nominalisation |               |
|--------------|-----------------|---------------|----------------|---------------|
|              | IELTS-rated     | Teacher-rated | IELTS-rated    | Teacher-rated |
| Thesis       | 45%             | 57%           | 0%             | 24%           |
| Argument 1   | 54%             | 58%           | 12%            | 21%           |
| Argument 2   | 59%             | 63%           | 10.5%          | 12%           |
| Argument 3   | 51%             | -             | 16%            | -             |
| Restatement  | 47%             | 58%           | 22%            | 31%           |
| Means        | 51.2%           | 59%           | 12.1%          | 22%           |

**Highly-rated IELTS essay**

| Stage                                             | Lexical items used in texts                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis<br><b>L. density</b><br>18/40<br>(45%)     | <u>Some people argue</u> that the <u>state</u> does not <u>have</u> the <u>right</u> to <u>make</u> <u>parents</u> <u>immunise</u> their <u>children</u> . <u>However</u> , I <u>feel</u> <u>the</u> <u>question</u> is not whether they should <u>immunise</u> but whether, as <u>members</u> of <u>society</u> , they <u>have</u> the <u>right</u> not to.                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>Nominalisation</b><br>0/18<br>(0%)             | Some people argue that the state does not have the right to make parents immunise their children. However, I feel the question is not whether they should immunise but whether, as members of society, they have the right not to.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Argument 1<br><b>L. density</b><br>25/46<br>(54%) | <u>Preventative medicine</u> has <u>proven</u> to be the <u>most</u> <u>effective</u> <u>way</u> of <u>reducing</u> the <u>incidence</u> of <u>fatal</u> <u>childhood</u> <u>diseases</u> . As a <u>result</u> of the <u>widespread</u> <u>practice</u> of <u>immunising</u> <u>young</u> <u>children</u> in our <u>society</u> , many <u>lives</u> have been <u>saved</u> and the <u>diseases</u> have been <u>reduced</u> to <u>almost</u> <u>zero</u> .                                                                                                                                                                                                                                                                                      |
| <b>Nominalisation</b><br>3/25<br>(12%)            | <u>Preventative medicine</u> has <u>proven</u> to be <u>the</u> <u>most</u> <u>effective</u> <u>way</u> of <u>reducing</u> the <u>incidence</u> of <u>fatal</u> <u>childhood</u> <u>diseases</u> . As a <u>result</u> of <u>the</u> <u>widespread</u> <u>practice</u> of <u>immunising</u> <u>young</u> <u>children</u> in our <u>society</u> , many lives have been saved and the diseases have been reduced to almost zero.                                                                                                                                                                                                                                                                                                                   |
| Argument 2<br><b>L. density</b><br>38/65<br>(59%) | In <u>previous</u> <u>centuries</u> <u>children</u> <u>died</u> from <u>ordinary</u> <u>illnesses</u> <u>such</u> <u>as</u> <u>influenza</u> and <u>tuberculosis</u> and because few people had <u>immunity</u> , the <u>diseases</u> <u>spread</u> <u>easily</u> . <u>Diseases</u> <u>such</u> <u>as</u> <u>dysentery</u> were the <u>result</u> of <u>poor</u> <u>hygiene</u> but these have <u>long</u> <u>been</u> <u>eradicated</u> since the <u>arrival</u> of <u>good</u> <u>sanitation</u> and <u>clean</u> <u>water</u> . <u>Nobody</u> would <u>suggest</u> that we should <u>reverse</u> this <u>good</u> <u>practice</u> <u>now</u> because <u>dysentery</u> has been <u>wiped</u> <u>out</u> .                                     |
| <b>Nominalisation</b><br>4/38<br>(10.5%)          | In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had <u>immunity</u> , the diseases spread easily. Diseases such as dysentery were the result of <u>poor</u> <u>hygiene</u> but these have long been eradicated since <u>the</u> <u>arrival</u> of <u>good</u> <u>sanitation</u> and <u>clean</u> <u>water</u> . Nobody would suggest that we should reverse this <u>good</u> <u>practice</u> now because dysentery has been wiped out.                                                                                                                                                                                                                                    |
| Argument 3<br><b>L. density</b><br>37/73<br>(51%) | <u>Serious</u> <u>diseases</u> <u>such</u> <u>as</u> <u>polio</u> and <u>smallpox</u> have also been <u>eradicated</u> through <u>national</u> <u>immunisation</u> <u>programmes</u> . In <u>consequence</u> , <u>children</u> not <u>immunised</u> are <u>far</u> <u>less</u> <u>at</u> <u>risk</u> in this <u>disease-free</u> <u>society</u> than they <u>otherwise</u> be. <u>Parents</u> <u>choosing</u> not to <u>immunise</u> are <u>relying</u> on the <u>fact</u> that the <u>diseases</u> have already been eradicated. If the <u>number</u> of <u>parents</u> <u>choosing</u> not to <u>immunise</u> <u>increased</u> , there would be a <u>similar</u> <u>increase</u> in the <u>risk</u> of the <u>diseases</u> <u>returning</u> . |

|                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Nominal<br/>isation</b><br>6/37<br>(16%)            | Serious diseases such as polio and smallpox have also been eradicated through <u>national immunisation programmes</u> . In consequence, <u>children not immunised</u> are far less at risk in this <u>disease-free society</u> than they otherwise be. <u>Parents choosing not to immunise</u> are relying on the fact that the diseases have already been eradicated. If <u>the number of parents choosing not to immunise</u> increased, there would be <u>a similar increase in the risk of the diseases returning</u> . |
| Restatement<br><br><b>L. density</b><br>23/49<br>(47%) | <u>Immunisation</u> is not an <u>issue</u> like <u>seatbelts</u> which <u>affect only</u> the <u>individual</u> . A <u>decision</u> not to <u>immunise</u> will <u>have widespread repercussions</u> for the <u>whole of society</u> and for this <u>reason</u> , I do not <u>believe</u> that <u>individuals</u> <u>have the right to stand aside</u> . In my <u>opinion</u> <u>immunisation</u> should be <u>obligatory</u> .                                                                                             |
| <b>Nominal<br/>isation</b><br>5/23<br>(22%)            | <u>Immunisation</u> is not an <u>issue</u> like <u>seatbelts</u> which <u>affect only</u> the <u>individual</u> . A <u>decision</u> not to <u>immunise</u> will have <u>widespread repercussions</u> for the whole of society, and for this reason, I do not <u>believe</u> that <u>individuals</u> <u>have the right to stand aside</u> . In my <u>opinion</u> <u>immunisation</u> should be <u>obligatory</u> .                                                                                                           |

#### Essay highly rated by teacher

| Stage                                              | Lexical items used                                                                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis<br><b>L. density</b><br>21/37<br>(57%)      | In the <u>past 20 years</u> , <u>globalisation</u> has <u>brought people closer</u> . It is the <u>process of opening up domestic economies to foreign competition and investment</u> (EAP-4, p.64). <u>Globalisation brings several negative impacts to the economies of developing countries</u> .                                                                                  |
| <b>Nominal<br/>isation</b><br>5/21<br>(24%)        | In the past 20 years, <u>globalisation</u> has brought people closer. It is <u>the process of opening up domestic economies to foreign competition and investment</u> (EAP-4, p.64). <u>Globalisation</u> brings <u>several negative impacts to the economies of developing countries</u> .                                                                                           |
| Argument 1<br><b>L. density</b><br>28/48<br>(58%)  | One of the <u>negative impacts of globalisation</u> is <u>tax evasion</u> . <u>Tax avoidance by some multinational companies costs the developing nations approximately \$160 billion every year</u> (Peake, 2011). Moreover, the <u>profits of the developing countries have extremely decreased</u> as a <u>result of the tax dodging by some multinational companies</u> .         |
| <b>Nominal<br/>isation</b><br>6/28<br>(21%)        | <u>One of the negative impacts of globalisation</u> is <u>tax evasion</u> . <u>Tax avoidance by some multinational companies</u> costs <u>the developing nations</u> approximately \$160 billion every year (Peake, 2011). Moreover, <u>the profits of the developing countries</u> have extremely decreased as a <u>result of the tax dodging by some multinational companies</u> .  |
| Argument 2<br><b>L. density</b><br>65/104<br>(63%) | <u>Secondly, globalisation leads to exploitation of human resources</u> . <u>Labourers working in multinational companies which are set up in developing countries such as Indonesia, China and India are generally underpaid</u> . The <u>employees in some multinational companies</u> have to <u>work overtime but receive low wages</u> . Furthermore, the <u>celebrities who</u> |

|                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|--------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                        | promote the goods are paid more than the workers who have worked very hard to produce the goods. For example, the celebrities get paid about \$1 million for advertising a product whereas the workers get only \$5 per day to produce it. Therefore, the number of poor people is increasing due to the exploitation of human resources.                                                                                                                                                                                                                                                                                                                                            |
| <b>Nominalisation</b><br>8/65<br>(12%)                 | Secondly, globalisation leads to exploitation of human resources. Labourers working in multinational companies which are set up in developing countries such as Indonesia, China and India are generally underpaid. The employees in some multinational companies have to work overtime but receive low wages. Furthermore, the celebrities who promote the goods are paid more than the workers who have worked very hard to produce them. For example, the celebrities get paid about \$1 million for advertising a product whereas the workers get only \$5 per day to produce it. Therefore, the number of poor people is increasing due to the exploitation of human resources. |
| Restatement<br><br><b>L. density</b><br>29/50<br>(58%) | In conclusion, the negative impacts of globalisation have affected the economies in developing countries, and they were very intense. Two of the most serious impacts are tax evasion and exploitation of human resources. Tax evasion by the multinational companies and exploitation of human resources have caused poverty in developing countries.                                                                                                                                                                                                                                                                                                                               |
| <b>Nominalisation</b><br>9/29<br>(31%)                 | In conclusion, the negative impacts of globalisation have affected the economies in developing countries, and they were very intense. Two of the most serious impacts are tax evasion and exploitation of human resources. Tax evasion by the multinational companies and exploitation of human resources have caused poverty in developing countries.                                                                                                                                                                                                                                                                                                                               |

Overall the two essays display the language of mature writers by means of high lexical density and a variety of grammatical metaphor (see Table 5.4 above). Based on Ure's 1971 study (cited in Ventola 1996), the high percentages of lexical density across the stages of genre indicate the comparability of the essays to an academic article. Moreover, grammatical metaphors are employed in all Argument stages of the essays to create "buried" reasoning (reasoning inside clauses) rather than "unpacked" reasoning (reasoning between clauses) (Martin 1991). When using buried reasoning, the message or information is packed into the participants, creating long and complex nominal groups using pre-modifier and/or post-modifier but very simple syntactic relations (Ravelli 1988). In unpacked reasoning, on the

other hand, longer clauses are needed to carry the same information because it is loaded in the process (verb) of a clause. For example, as seen in the Argument 1 below, the unpacking of the nominal group (*tax avoidance by some multinational companies*) requires a whole clause (*There are some companies which operate around the world that avoid paying taxes*) to deliver the same information. This example corroborates Halliday (2001, p. 182) who argues that spoken language is organised around the clause, in the sense that most of the experiential content is laid down in the transitivity system. By the same token, written language is organised around the nominal group.”

Some samples of buried reasoning in the Argument stages of the essays are presented below:

(in Argument 1)

|                           |                                                                                                                                                                                                        |
|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Buried reasoning</b>   | <u>Tax avoidance by some multinational companies</u> costs <u>the developing nations</u> approximately \$160 billion every year.                                                                       |
| <b>Unpacked reasoning</b> | <u>There are some companies which operate around the world that avoid paying taxes.</u> Because they avoid paying taxes, <u>the third world countries</u> lose approximately \$160 billion every year. |

(in Argument 2)

|                           |                                                                                                                                                                                                                   |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Buried reasoning</b>   | Secondly, <u>globalisation</u> leads to <u>exploitation of human resources</u>                                                                                                                                    |
| <b>Unpacked reasoning</b> | Secondly, <u>the process of opening up domestic economies to foreign competition and investment</u> makes overseas companies <u>pay workers low wages and force them to work long hours in unsafe conditions.</u> |

(in Argument 3)

|                           |                                                                                                                                                                              |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Buried reasoning</b>   | Serious diseases such as polio and smallpox have also been eradicated through <u>national immunisation programmes.</u>                                                       |
| <b>Unpacked reasoning</b> | Serious diseases such as polio and smallpox have also been eradicated because <u>vaccine is used to immunize children around the nations against those serious diseases.</u> |

Using buried rationale such as the above makes argument sound more plausible as the interpretations are presented as a fact. This strategy of reasoning using incongruent realisations (metaphors) predominate in adults' academic written texts and is favoured by the discourse community (Halliday 1985b, Martin 1985).

#### 5.1.4 Analysis of the textual function of text: the Theme system

As explained previously, (see Chapter Four, Section 4.6.1.2), analysis of text structure is concerned with the organisation of messages in a way that makes them fit in smoothly with other parts. With written text lying at the reflective end of the mode continuum (see Eggins 1994 , 2004, Martin 1984), there is an expectation that it be carefully organised. The careful organisation of the text by means of choices of macroThemes and hyperThemes can orient the readers to what the text is about (Martin and Rose 2003; 2007). Similarly, by means of Theme-Rheme structures, information is expected to develop gradually as the text unfolds. Well-signposted progression of Themes-Rhemes will keep the reader on track when reading the text. The following section illustrates macroThemes and hyperThemes, followed by more detailed information on the Theme-Rheme structures in the teacher-rated essay.

#### **5.1.4.1 macroTheme and hyperTheme**

The system of Theme, which is a textual resource at clause rank for representing clause as a message, plays a significant part in constructing the flow and organisation of information in a text (Halliday 1994, Martin, Matthiessen et al. 1997). At higher levels of discourse, beyond the clause rank, the Theme development creates a bigger wave of information flow known as macroTheme and hyperTheme (Martin and Rose 2007). While the macroTheme would orient the reader(s) to what the whole essay is about, the hyperThemes would inform the reader(s) on the content of each Argument stage of the essay (Butt, Fahey et al. 2000).

The information is systematically arranged in the student essay. In the Thesis stage, the macroTheme identifies negative impacts of globalisation on the economies of developing countries. These provide previews for the readers and predict the organisation of the following Argument stages (Coffin 1997, Rose and Martin 2012). The first hyperTheme links back to the economic issue as a result of tax evasion, and the second hyperTheme to the economic issue as a result of exploitation of human resources. These points are then elaborated in turn. In the Restatement stage, the argument announced in the macroTheme is reiterated by using some key words: "negative impacts", "economies", "tax evasion" and "exploitation of human resources".

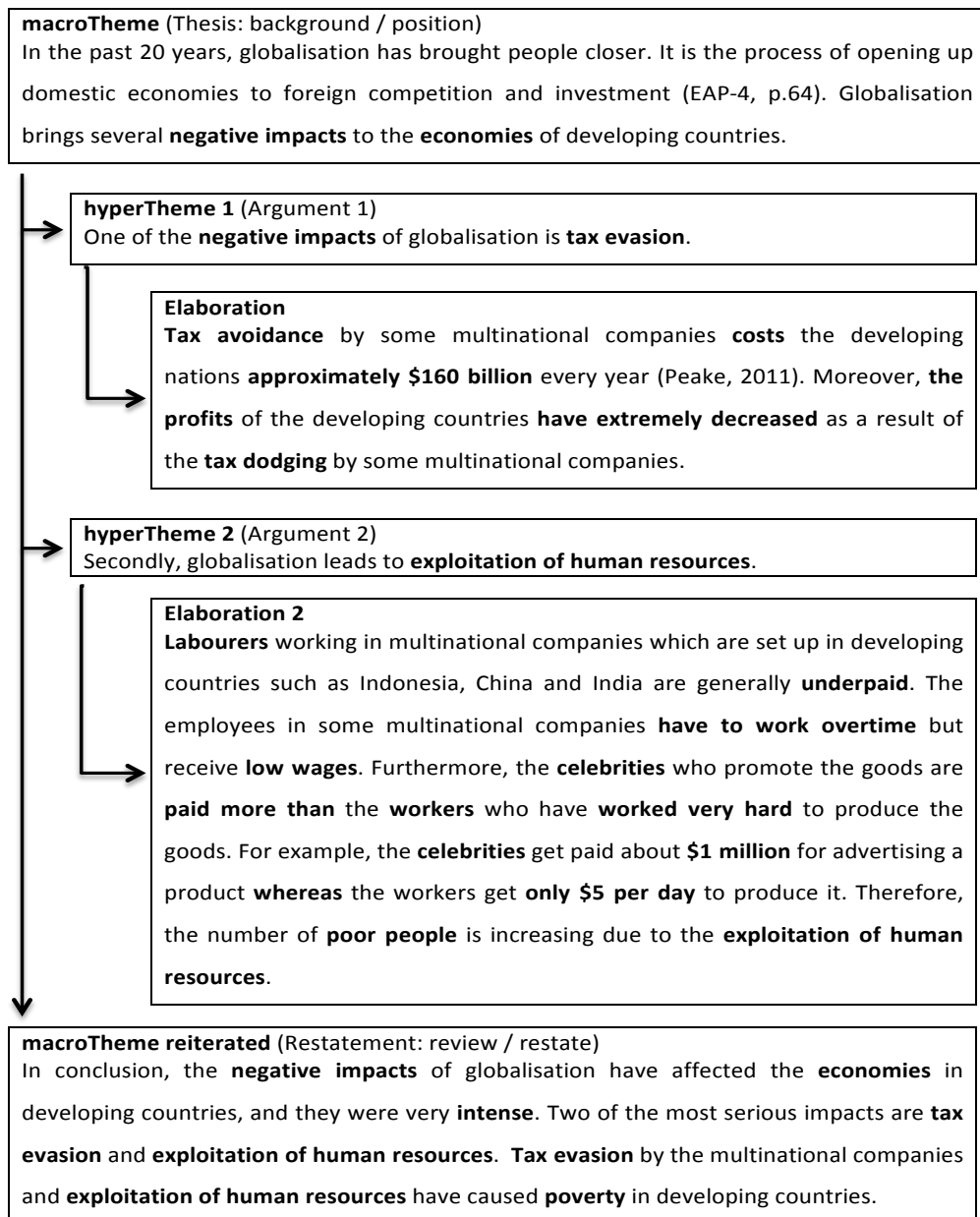


Figure 5.1 macroTheme and hyperTheme development in teacher-rated essay

The next section further illustrates Theme-Rheme structures of each stage of the teacher-rated essay.

#### 5.1.4.2 Theme-Rheme structures

The Theme system is concerned with the structural organisation of text, and the kinds of Themes used indicate different approaches to text organisation

(Schleppegrell 2004). Halliday (2001; 2004) argues that academic, particularly scientific, texts are thematised around abstract and dense noun groups while everyday texts typically thematise pronominal items (e.g. *they*, *the workers* or *my parents*). Based on this argument, it can be assumed that the essay is highly written as the clause Themes reveal that the essay is, overall, organised using predominantly abstract Themes (Martin 2001). This means that the student organises his/her paragraphs by predominantly employing abstraction around *the impacts of globalisation* and *the developing countries*. In order to do this, the student makes use of abstract Themes such as *negative impacts globalisation*, *tax avoidance by some multinational companies*, *the profits of the developing countries*, and *serious impacts*, (which are also nominalisations). Even though some ‘spoken Themes’ such as *they*, *the celebrities* and *the workers* are used in the text, they all function to elaborate the new information given (*the exploitation of human resources*) in the Rheme of the second Argument stage (the Theme-Rheme relations of clauses in the second Argument stage can be seen in Figure 5.4). Thus, it can be said that the student strategically employs the choice of Themes to improve the text’s focus (Christie and Rothery 1989), which in this case is the adverse effect of globalisation on third-world countries. As we shall see, the use of variety Thematic progression (Theme reiteration, zigzag, and multiple Rheme patterns) demonstrates the rhetorical skill of the writer to craftily present his/her knowledge about the issue, which Bereiter and Scardamalia (1987) referred to as ‘knowledge transforming’.

The analysis of coding for Theme-Rheme structures is as follows: **Topical Theme** is boxed, Interpersonal Theme is underlined, *Textual Theme* is italicised and the **marked** status of both Topical theme and Clause as Theme is in bold. The ellipsed

participants and processes are put in brackets preceded by a caret (^), and the *relative pronouns* which function as confluents of textual and topical theme are italicised and boxed.

#### The Thesis stage of teacher-rated essay

| Phase        | Clause | Theme                        | Rheme                                                                                                |
|--------------|--------|------------------------------|------------------------------------------------------------------------------------------------------|
| 'background' | 1      | <b>In the past 20 years,</b> | globalisation has brought people closer.                                                             |
|              | 2      | <i>It</i>                    | is the process of opening up domestic economies to foreign competition and investment (EAP-4, p.64). |
| 'position'   | 3      | <i>Globalisation</i>         | brings several negative impacts to the economies of developing countries.                            |

To attract attention, the Thesis stage is started off with a marked topical Theme realised in adverbial (circumstantial element of time), *In the past 20 years*, to construe temporal succession. Although at first glance the marked Theme used seems to be more appropriate for historical genre, it actually serves its function very well as the background information to draw the readers' attention before leading them to the topic or field to be discussed. The first key word relevant to the essay prompt, *globalisation*, is then put immediately as the first Rheme. The key word is then used as a Theme in the next two clauses. The thematic progression of ideas in the Thesis stage is seen as realised through a multiple-Rheme pattern as presented in Figure 5.2, below:

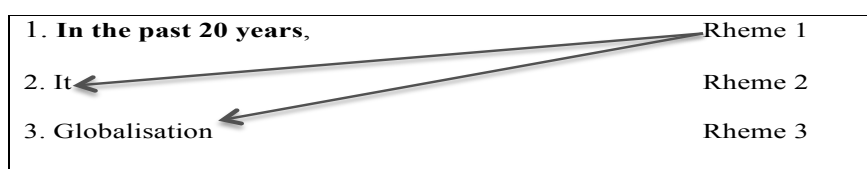


Figure 5.2 Teacher-rated essay: Thematic progression in the Thesis stage

### The first argument stage of teacher-rated essay

| Phase         | Clause | Theme                                             | Rheme                                                                                    |
|---------------|--------|---------------------------------------------------|------------------------------------------------------------------------------------------|
| 'topic'       | 1      | One of the negative impacts of globalisation      | is tax evasion.                                                                          |
| 'elaboration' | 2      | Tax avoidance by some multinational companies     | costs the developing nations approximately \$160 billion every year (Peake, 2011).       |
|               | 3      | Moreover, the profits of the developing countries | have extremely decreased as a result of the tax dodging by some multinational companies. |

After indicating the position of the writer in the Thesis stage by using two key words (*impacts* and *economies*), the first Argument stage can then take the negative impacts of globalisation as the first topical Theme in the topic phase. Also, as expected in the previous stage (macroTheme), the negative impacts of globalisation are discussed in relation to economic issues. The economic issue is realised as “tax evasion” in the Rheme of the first clause. This Rheme is then used as the topical Theme in the first elaboration phase by using a synonym (*tax evasion* – *tax avoidance*). Subsequently, the Rheme of the previous clause (*costs the developing nations*) is picked up as the Theme in the next clause. The use of a conjunctive adjunct *moreover* functions to signal the textual organisation, in that it serves to link clauses together and at the same time signals additional information to the previous clause (Halliday 1994, Thompson 1996). The thematic progression of ideas in the Argument 1 stage is realised through zigzag Theme pattern in that information introduced in the previous sentence is picked up in the Theme of the following sentence (see Figure 5.3).

|                                                  |         |
|--------------------------------------------------|---------|
| 1. One of the negative impacts of globalisation  | Rheme 1 |
| 2. Tax avoidance by some multinational companies | Rheme 2 |
| 3. the profits of the developing countries       | Rheme 3 |

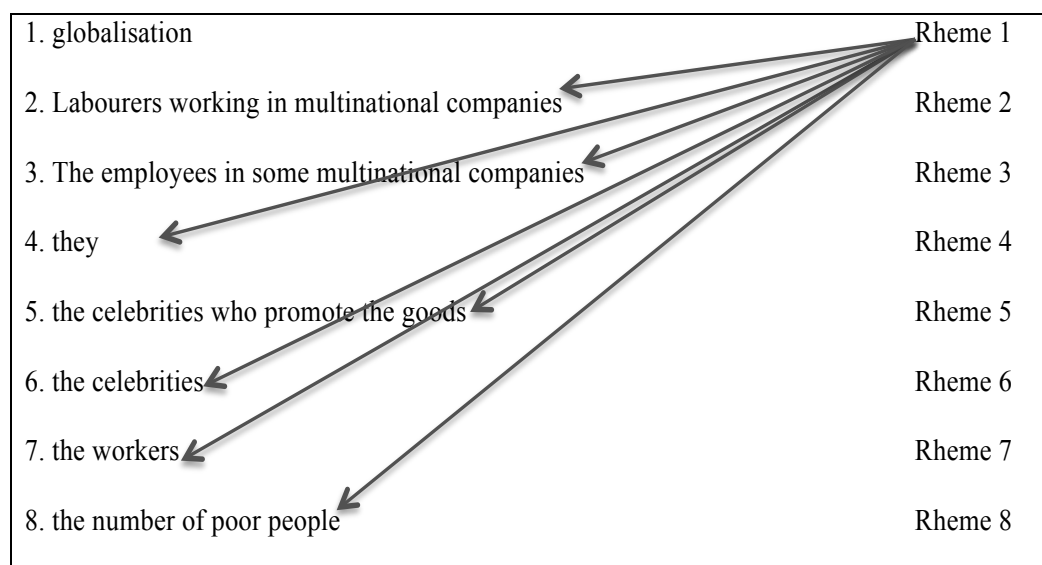
Figure 5.3 Teacher-rated essay: Thematic progression in the Argument 1 stage

### The second argument stage

| Phase              | Cl | Theme                                                                                                                        | Rheme                                                                     |
|--------------------|----|------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| <i>topic</i>       | 1  | <i>Secondly, globalisation</i>                                                                                               | leads to exploitation of human resources.                                 |
| <i>elaboration</i> | 2  | Labourers working in multinational companies [[which are set up in developing countries such as Indonesia, China and India]] | are generally underpaid.                                                  |
|                    | 3  | The employees in some multinational companies                                                                                | have to work overtime                                                     |
|                    | 4  | but (^they)                                                                                                                  | receive low wages.                                                        |
|                    | 5  | Furthermore, the celebrities who promote the goods                                                                           | are paid more than the workers who have worked very hard to produce them. |
|                    | 6  | For example, the celebrities                                                                                                 | get paid about \$1 million for advertising a product                      |
|                    | 7  | whereas the workers                                                                                                          | get only \$5 per day to produce it.                                       |
|                    | 8  | Therefore, the number of poor people                                                                                         | is increasing due to the exploitation of human resources.                 |

The second argument stage begins by taking the Theme of ‘position’ phase (*globalisation*) in the macroTheme stage as in the ‘topic’. The new information (Rheme) of the topic in clause 1 is used as the Themes in the subsequent clauses: *labourers working in multinational companies* as in clause 2, *the employees* as in clause 3 and ellipsis subject *they* as in clause 4. In clauses 5, 6 and 7, the Themes are more focused on the *exploitation of human resources* (the Rheme in the topic phase), by showing and illustrating disparities of income between celebrities who endorse the products and the workers who produce them. The last clause of the stage started with a textual Theme (*therefore*) which functions as a concluding remark to

signal that the writer is about the close his/her case. Again, the writers' use of topical Theme *the number of poor people* links back directly to the Rheme of clause 1. Overall, the progression of ideas in the second argument stage can be seen as realised through multiple-Rheme pattern as illustrated in Figure 5.4.

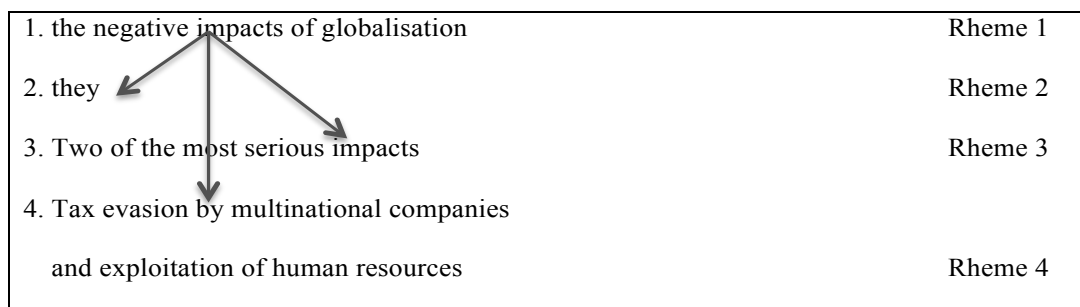


**Figure 5.4 Teacher-rated essay: Thematic progression in the Argument 2 stage**

### The Restatement stage of teacher-rated essay

| Phase     | Clause | Theme                                                                                                                     | Rheme                                                |
|-----------|--------|---------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------|
| 'review'  | 1      | <i>In conclusion, the</i><br><i>negative impacts of</i><br><i>globalisation</i>                                           | have affected the economies in developing countries, |
|           | 2      | <i>and they</i>                                                                                                           | are very intense.                                    |
| 'restate' | 3      | <i>Two of the most</i><br><i>serious impacts</i>                                                                          | are tax evasion and exploitation of human resources. |
|           | 4      | <i>Tax evasion by</i><br><i>multinational</i><br><i>companies and</i><br><i>exploitation of human</i><br><i>resources</i> | have caused poverty in developing countries          |

The stage opens with a textual Theme (*In conclusion*), which acts as a marker to signal closure or the summation of information (Christie 2002, Halliday 1985a). The key words mentioned in the Thesis are affirmed in the *review* phase using the Theme-Rheme pattern (*negative impacts of globalisation-economies of developing countries*). In the *restate* phase, the writer lays the case by restating the two negative impacts discussed in the body paragraphs using a whole-part lexical relationship *economy – poverty*, known as meronymy. The thematic progression using taxonomic relation in the teacher-rated essay such as repetition (*negative impacts – most serious impacts*), synonymy (*exploitation of human resources – underpaid – overtime – low wages*) and meronymy (*economy – poverty*) contributes to the cohesiveness of the text (mode) and at the same time helps construe the topic (field) of the essay (Martin 1997, Martin and Rose 2003; 2007).



**Figure 5.5 Thematic progression in the Restatement stage**

### **5.1.5. Analysis of the interpersonal function of text: the Appraisal system**

As previously discussed (see Chapter Three, Section 3.3.4), the system of Appraisal provides a systematic tool to help with exploring and explaining how language is used to evaluate, to adopt stance and manage relationship, and to scale up/down the evaluative positioning (White 2006; 2012). These three main functions are respectively realised by the three Appraisal resources of ATTITUDE, ENGAGEMENT and GRADUATION. The system of ATTITUDE covers evaluations which source from emotive feelings (Affect), evaluations of things or phenomena (Appreciation), and evaluations of people's behaviours (Judgement). The literature has shown that quality academic writing is characterised by the prevalence of ATTITUDE expressed as Appreciation over Affect and Judgement (see, for example, Derewianka 2007, Hood 2004, S. H. Lee 2006; 2007, Nakamura 2009 among others). Similarly, the two essays were analysed to discern how students evaluated the issues under discussion, how they 'turned up or turned down' the evaluations to persuade their readers, and how they engaged with other voices or perspectives around the issues (Coffin and Hewings 2004).

Hood (2004) argues that Appreciation is a means by which writers can appear more objective in terms of their value positions. The analysis of the two essays reveals the prevalent employment of Appreciation resource in the evaluations made. In the IELTS-rated essay, positive Appreciation, such as *most effective way* and *good practice*, is predominantly used when evaluating the field of discussion (immunisation) because the writer's position is to argue for national immunisation program. Thus, the predominance of positive Appreciation in the text is not unforeseen. In the teacher-rated essay, however, negative Appreciation, such as *negative impacts* and *serious impacts*, is more pervasively used due to the fact that the writer positions him/herself against globalisation. In addition, implicit (invoked in Appraisal term) Appreciation elements, such as *As a result of the widespread practice of immunising young children in our society, many lives have been saved...* and *globalisation leads to exploitation of human resources* are predominant in the two essays. As seen in the two examples above, there is no “attitudinal lexis that tells us directly how to feel, but the selection of ideational meanings is enough to invoke evaluation” (Martin and White 2005, p. 62). The prevalence of Appreciation and the absence of Affect confirm the high quality of both essays, and at the same time reverberate the findings of the studies previously discussed (see, for example, Derewianka 2007, Hood 2004, S. H. Lee 2006, Nakamura 2009).

### IELTS-rated essay

|                                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis                                                                                                                                                                                                                             |
| Some people argue that the state does not have the right to make parents immunise their children. However, I feel the question is not whether they should immunise but whether, as members of society, they have the right not to. |
| Argument 1                                                                                                                                                                                                                         |
| Preventative medicine has proved to be the most effective way of reducing the incidence of fatal childhood diseases. As a result of the widespread practice of                                                                     |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| immunising young children in our society, many lives have been saved and the diseases have been reduced to almost zero.                                                                                                                                                                                                                                                                                                                                                               |
| Argument 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had immunity, the diseases spread easily. Diseases such as dysentery were the result of poor hygiene but these have long been eradicated since the arrival of good sanitation and clean water. Nobody would suggest that we should reverse this good practice now because dysentery has been wiped out.                                                         |
| Argument 3                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Serious diseases such as polio and smallpox have also been eradicated through national immunisation programmes. In consequence, children not immunised are far less at risk in this disease-free society than they would otherwise be. Parents choosing not to immunise are relying on the fact that the diseases have already been eradicated. If the number of parents choosing not to immunise increased, there would be a similar increase in the risk of the diseases returning. |
| Restatement                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Immunisation is not an issue like seatbelts which affect only the individual. A decision not to immunise will have widespread repercussions for the whole of society and for this reason, I do not believe that individuals have the right to stand aside. In my opinion immunisation should be obligatory.                                                                                                                                                                           |

### Teacher-rated essay

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thesis                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| In the past 20 years, globalisation has brought people closer. It is the process of opening up domestic economies to foreign competition and investment (EAP-4, p.64). Globalisation brings several negative impacts to the economies of developing countries.                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Argument 1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| One of the negative impacts of globalisation is tax evasion. Tax avoidance by some multinational companies costs the developing nations approximately \$160 billion every year (Peake, 2011). Moreover, the profits of the developing countries have extremely decreased as a result of the tax dodging by some multinational companies.                                                                                                                                                                                                                                                                                                                                                  |
| Argument 2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Secondly, globalisation leads to exploitation of human resources. Labourers working in multinational companies which are set up in developing countries such as Indonesia, China and India are generally underpaid. The employees in some multinational companies have to work overtime but receive low wages. Furthermore, the celebrities who promote the goods are paid more than the workers who have worked very hard to produce the goods. For example, the celebrities get paid about \$1 million for advertising a product whereas the workers get only \$5 per day to produce it. Therefore, the number of poor people is increasing due to the exploitation of human resources. |
| Restatement                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| In conclusion, the negative impacts of globalisation have affected the economies in developing countries, and they were very intense. Two of the most serious impacts are tax evasion and exploitation of human resources. Tax evasion by the multinational companies and exploitation of human resources have caused poverty in developing countries.                                                                                                                                                                                                                                                                                                                                    |

Distribution of Appraisal resources counts shows that the two essays are comparable to each other. In terms of ATTITUDE resources, Appreciation, with the ratio of 34%, is the most prevalently used resource in the two essays while Affect is barely available in the text.

**Table 5.5 Distribution of Appraisal resources in the two essays**

| IELTS-rated essay |     |       |      |     | Teacher-rated essay |     |      |      |     |
|-------------------|-----|-------|------|-----|---------------------|-----|------|------|-----|
| Aff               | App | Judg  | Grad | Eng | Aff                 | App | Judg | Grad | Eng |
| 1.5%              | 34% | 12.5% | 27%  | 25% | -                   | 34% | 15%  | 30%  | 21% |

The two essays are also compatible in terms of GRADUATION resources with the ratio of 27% and 30% of the Appraisal items respectively. The GRADUATION resources frequently function to specify appraised entities, demonstrating the writers' expertise on the topics under discussion (fields) and at the same time, lead to more convincing arguments.

*In previous centuries children died from **ordinary** illnesses such as influenza and tuberculosis... (IELTS-rated essay, Argument 2)*

***Serious** diseases such as polio and smallpox have also been eradicated through national immunisation programmes. (IELTS-rated essay, Argument 3)*

*Tax avoidance by some multinational companies costs the developing nations **approximately \$160 billion** every year. (Teacher-rated essay, Argument 1)*

*For example, the celebrities get paid **about \$1 million** for advertising a product whereas the workers get **only \$5** per day to produce it. (Teacher-rated essay, Argument 2)*

Another compatibility is the ratio of ENGAGEMENT resources in the two essays (25% for the IELTS-marked essay and 21% for the teacher-rated essay), which indicates a high degree of dialogic alternatives. In the Restatement stage of the IELTS-rated essay, for example, the coding the statement *In my opinion immunisation should be*

*obligatory* (heteroglossic/contract/pronounce in Appraisal term) suggests that the student positions him/herself on one of the other voices. The pervasive employment of dialogic contraction e.g., *however*, *but* (counter), *not* (deny), *has proven* and *have caused* (pronounce) across the texts indicates that the arguments are mostly built to fend off positions being held. However, the presence of dialogic expansion features shows the willingness of the writer to opt for other possibilities and at the same time functions to show neutrality of position in the arguments made. Some examples are provided below:

*Some people argue that the state does not have the right to make parents immunise their children. (IELTS-rated essay, Thesis)*

*However, I feel the question is not whether they should immunise but whether, as members of society, they have the right not to. (IELTS-rated essay, Thesis)*

*The employees in some multinational companies have to work overtime but receive low wages. (Teacher-rated essay, Argument 2)*

In summary, the linguistic analysis of the two successful essays revealed the following features indicative of quality essays. At the level of genre, quality essays are well organised in terms of stages and phases to achieve the purpose of essay writing. At the register level, mastery or knowledge of the topic being discussed is evident in the selection of nominal groups with abstract heads such as *globalisation*, *exploitation of human resources* and *immunisation*, and the sentences are lexically dense, in that the information is packed into noun groups. These are clearly shown by the analyses of Clause complex and lexical density/nominalisation. The appraisal analysis of the two quality essays shows that phenomena or issues are evaluated

using less personal language, indicated by the prevalence of Appreciation over Affect or Judgement. The progression of ideas is clear and cohesive, creating coherence of text. This was gained from the Theme analysis of students' essays. In addition, teachers nominated grammatical accuracy as features of successful essays.

These features were adopted in the new rubric by summarising them into three main features. The first covers the purpose of the essay and stages and phases used to achieve its purpose. The second consists of three main areas representing the three functions of language (to express ideas, to interact with others and to compose message). The last part of the new rubric is concerned with accuracy of language use such as correctness in terms of grammar, word spelling and punctuation.

The next section of the thesis is concerned with the applications of the features of the successful texts on the design of the new rubric and intervention.

## **5.2 Establishing an SFL-informed rubric**

The design of the new rubric, was particularly informed by the work of Rose (2005). In order to address the range of language resources that students used in their writing tasks, there needed to be some concise descriptions of the score traits. This would give clearer pictures of the scores the teachers should assign to each criterion in the new rubric. Adapted from Rose (2005, p. 1), the score traits of the SFL-informed rubric were as follows:

**Table 5.6 Score traits of the SFL-informed rubric (Rose 2005, p. 1)**

| <b>Marking Rubric</b>                                                                                                                                                                                                                                                                  |                                                                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|
| This SFL-informed marking rubric consists of 13 criteria which are used at the levels of genre, register, general academic essay features. These 13 criteria assess all the language resources that EAP-4 students use in their writing tasks. Each criterion is given a score of 0-3. |                                                                                                |
| 3                                                                                                                                                                                                                                                                                      | Student's writing meets a criterion at the highest level expected as explicated in the rubric. |
| 2                                                                                                                                                                                                                                                                                      | Quite strong evidence but not at the top standard                                              |
| 1                                                                                                                                                                                                                                                                                      | There is evidence but it is weak                                                               |
| 0                                                                                                                                                                                                                                                                                      | There is no evidence of the criterion                                                          |

As discussed in the previous chapter, the SFL-informed rubric was chosen because it enabled the description of texts at a number of levels (discourse, clause and word) and the intervention rendered visible linguistic expressions of what was valued and therefore, enabled an explicit shared understanding among markers. Adapted from Rose's assessment criteria (2005, p. 2), the draft of the SFL-informed rubric in this study was designed as a three-layer marking rubric. This reflected the SFL model of context-text relations discussed in Chapter Three (see, for example, Halliday and Hasan 1985, Halliday and Matthiessen 2004, Martin 1993). The top layer of the rubric (related to the context of culture) assessed student essays in relation to genre. Thus, student essays were marked in terms of the overall communicative purpose, the stages and phases (as seen in Figure 3.3 and Figure 3.4 in Chapter Three, Section 3.3.1) used to achieve the purpose of the writing task. With respect to the **Thesis** stage of the argument, for example, assessment was made in relation (but not limited) to how students were approaching the topic i.e., the effects of globalisation on developing countries, and the preview of position through which the arguments would be developed. With respect to the **Argument** stage of the argument, the essays were assessed on how the main topics/themes were developed into a series of well-organised paragraphs. In the **Restatement** stage of the argument, assessment

was made of how students distilled the points that they had made and highlighted their significance. The sample of SFL-informed criteria on the level of genre is presented below:

**Table 5.7 Genre**

| <b>Genre: social purpose and stages developed</b> |                                                                                                                                                                                        |   |   |     |
|---------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|-----|
| <b>Thesis</b>                                     | A statement that immediately orientates the reader to the topic of the essay and how the writer will be approaching it. This might include some brief background to key concepts.      | 0 | 1 | 2 3 |
|                                                   | The preview of position and main themes through which this position/argument will develop in the first paragraph.                                                                      | 0 | 1 | 2 3 |
| <b>Arguments</b>                                  | Development of main themes and their relationships in a series of well organised paragraphs which follow the order given in the first paragraph (are logically and clearly organised). | 0 | 1 | 2 3 |
| <b>Restatement</b>                                | Conclusion with overall interpretation, summarising the themes of the topic and highlighting their significance in light of the approach taken in the subject.                         | 0 | 1 | 2 3 |

The middle layer (related to context of situation) offered criteria for the essays in terms of register variables. As seen in Table 5.8, the marking criteria in the middle layer included: the content of the essay which identifies the negative/positive impact of globalisation (**field**), the formality and word choices used to evaluate, persuade or engage readers (**tenor**), and the written features used to create cohesiveness (**mode**).

**Table 5.8 Register**

| <b>Register variables</b>                                                                                                   |   |   |   |   |
|-----------------------------------------------------------------------------------------------------------------------------|---|---|---|---|
| <b>Topic research and content (field)</b>                                                                                   |   |   |   |   |
| Identifies and explores negative/positive impacts of globalisation                                                          | 0 | 1 | 2 | 3 |
| Uses information from prescribed text, uses evidence and examples to support argumentation                                  | 0 | 1 | 2 | 3 |
| Answers the question                                                                                                        | 0 | 1 | 2 | 3 |
| <b>Communication roles (tenor)</b>                                                                                          |   |   |   |   |
| Formality (academic tone)                                                                                                   | 0 | 1 | 2 | 3 |
| Appraisal (word choices writers use to engage, persuade and evaluate)                                                       | 0 | 1 | 2 | 3 |
| <b>Written features (mode)</b>                                                                                              |   |   |   |   |
| Textual cohesion: sentence and paragraph construction and relationship, use of conjunctions, synonyms and pronoun reference | 0 | 1 | 2 | 3 |

The lower layer (related to text) assessed them in relation to the linguistic features such as: grammatical accuracy such as tenses and Subject-Verb Agreement, spelling, punctuation, and word limit (see Table 5.9). When redesigning the new rubric, I also recognised what the teachers would be familiar and comfortable with. For that reason, I incorporated some elements from the old rubric to make the teachers feel more comfortable with using the new rubric in their practices. Therefore, for example, Subject-Verb Agreement, which was supposed to be part of the register variables of **mode**, was treated separately and included in the criteria of Accuracy.

**Table 5.9 Accuracy**

| <b>Grammatical and graphic features</b>                     |   |   |   |   |
|-------------------------------------------------------------|---|---|---|---|
| Spelling and punctuation                                    | 0 | 1 | 2 | 3 |
| Grammatical accuracy (i.e., tenses, subject-verb agreement) | 0 | 1 | 2 | 3 |
| Adherence to word limit requirement                         | 0 | 1 | 2 | 3 |

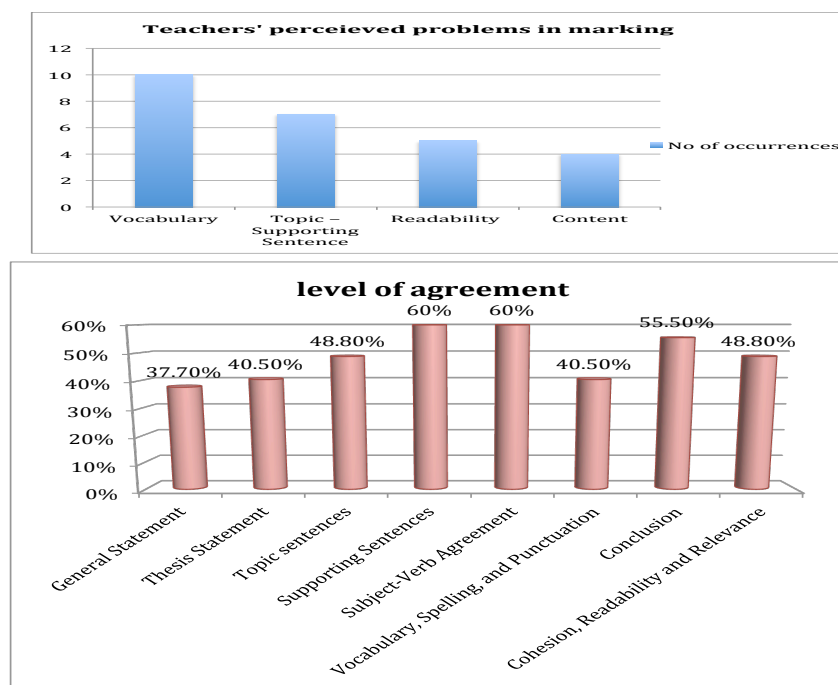
In addition to being informed by SFL theory, the criteria for the revised rubric were selected because of several reasons: they were items teachers were most inconsistent with; they were features that markers expressed most anxiety about; and they were selected because successful essays feature these choices (the analysis of a successful text is detailed in the next chapter). For example, *Grammatical accuracy* as a criterion was included because the analysis of pre-training interviews revealed some teachers' disapproval of using Subject-Verb (S-V) as a criterion of its own. The teachers remarked:

Let's see. When it goes to Subject-Verb Agreement, I don't think this is supposed to be one of the criteria because when they reach this kind of level they can do this academic style of English, most of them do not have problems with Subject-Verb Agreement. The only problem that they might have will be the structure (Jelita).

Grammar should be here, grammar or structure. So it's not simply about Subject-Verb Agreement (Yusuf)

Yeah, general grammar. Subject-Verb Agreement is important but we can break it down into other parts of grammar as well (Whitney).

Instead, the participants (e.g., Jelita, Yusuf and Whitney) suggested that it be changed to general grammar. This was because they perceived S-V Agreement as one of the features of grammar points, and at an intermediate-level of study students' grammatical competencies should not be assessed based only on S-V Agreement. The participants claimed not to find any serious problems with their students' use of S-V Agreement in their essays. This proposition was affirmed by the result of the pre-training inter-rater consistency counting (see Figure 5.6), which shows that S-V Agreement was one of the features with the highest consistency rate (60%).



**Figure 5.6 Teachers' perceived problems and level of agreement in terms of criteria used in the existing marking rubric**

## 5.3 Training teachers

### 5.3.1 The training cycle

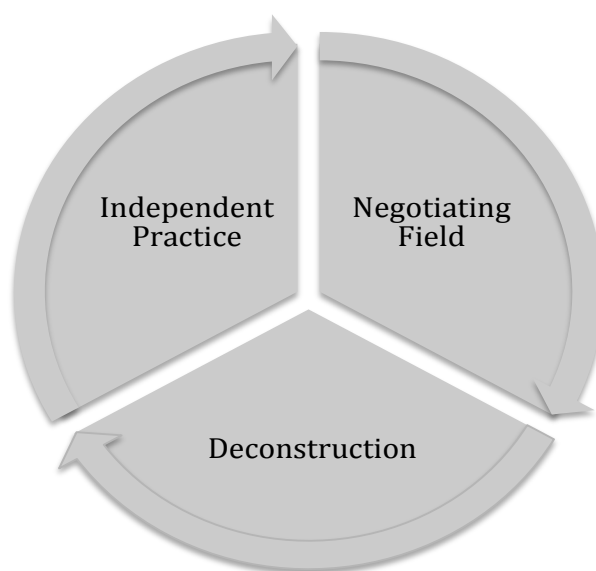
The intervention was conducted through four sessions of rater training (see Table 5.10 below). Each session ran once a week and lasted for two hours on weekdays or weekends. The teacher training was conducted in the classroom provided by the language school. The training program was planned as follows:

**Table 5. 10 Activities conducted during the intervention program**

| Session | Length  | Content                                                                                                                                                                                                                                                                                                        |
|---------|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| One     | 2 hours | Introduction to SFL: metafunctions and its relations to context upward and text downward<br><br>Practice: Group activities on field, tenor and mode                                                                                                                                                            |
| Two     | 2 hours | Review: Concept of genre and register<br><br>Practice: <ul style="list-style-type: none"><li>• Contextual analyses of texts</li><li>• Identifying features of essays in terms of content-carrying words, clause formation, thematic patterns between sentences and lexical cohesion at the paragraph</li></ul> |
| Three   | 2 hours | Introduction to Appraisal System<br><br>Practice: identifying Appraisal resources in a text                                                                                                                                                                                                                    |
| Four    | 2 hours | Introduction to SFL-informed rubric<br><br>Practice: identifying criteria on the level of genre, register and language.                                                                                                                                                                                        |

In all four sessions of the intervention program, the delivery methods of the training materials were informed by Rothery's (1996) Teaching and Learning Cycle. As the cycle is not considered as rigid, modifications of model can be made to suit the needs of students or trainees (Johns 1997). The students can enter at any stage of the cycle, skipping or omitting stages if they do not need them. In this study, the cycle also needed to be adapted in order to suit the needs of the students, who in this case,

were six EAP teachers. As explicated above, the purpose of the study was to enhance marking consistency by introducing teachers to SFL description of academic essay and SFL-informed rubric through an intervention program. Therefore, the focus of the intervention was on the explicit teaching of SFL concepts of genre, register and language functions followed by practice on related tasks. However, since the participants in this study were all experienced EAP teachers, practice in jointly constructing text in related genre was not necessary. Therefore, this study used the teaching cycle presented in Figure 5.7.



**Figure 5.7 The teaching cycle (adapted from Rothery 1996)**

Each training session started with researcher-led field-building activities to build teachers' knowledge (Negotiating Field). As informed above, SFL was a relatively new approach for most of the teachers. Thus, the first stage of the training cycle was of great importance, particularly during Session one and Session three. This was because the topics or fields discussed in these sessions (the relationship between context and Appraisal system) were relatively new domains for most of the teachers.

For example, after explaining the register variables in Session one, I showed the teachers how language choices varied from one situation to another. On the slide I provided an event or scenario as follows:

- **Event:** *As they sit watching television, a sixteen year-old girl is discussing with her mother what to buy her father for his birthday.*
- *Here the field is 'buying a gift'. The tenor is mother and daughter. The mode is face-to-face.*

After that, I provided other variations of the scenario such as the one seen below, and asked them questions in terms of Field, Tenor and Mode of the text.

- *Variation 2: a sixteen year-old girl is chatting with her boyfriend about what to buy his father for his birthday.*

The second stage of the training cycle was Deconstruction. In this stage, I worked with the participants in class to practise on the sample tasks given. During the training about context of culture and context of situation in Session two, for example, the teachers were asked to tease out or deconstruct the sample essays given based on each element of genre and register. Then the participants were asked to apply what they had learned by doing the related tasks given, starting from pair work and ending with whole-class work to discuss what had been done. In this stage of the cycle the teachers were asked in pairs to identify and mark the features of a sample argument essay and encouraged to discuss among themselves the reasons for the score(s) allocated.

In the final stage, the teachers were asked to work individually (Independent Practice). For example, in Session three, the teachers were asked to identify the

Appraisal resources found in the text after doing similar exercises in pairs. Similarly, in the last stage of the training cycle in Session four the teachers were asked to mark a sample argument essay by themselves. However, after having finished with their individual marking, the scores were analysed and discussed as a whole-class activity to reduce any misinterpretation of each criterion used. Before ending each session of the training, the teachers were given time for reflection so that they could assess and give meaning to their experience (Mezirow 1991); to reflect on their task doing performances, on whether there was something needed to be explained more clearly to have resulted in a better outcome.

### **5.3.2 Training sessions and activities**

#### **Session One**

Training Session One aimed at enabling teachers to see how the language use is dependent on the contexts of culture and the context of situation. Since most of the teachers had never learnt about functional grammar, the training was started with an introduction to SFL and its view of language compared to the traditional approach. I pointed out that while traditional approach views the language as a list of rules to be acquired so as to be able to use it correctly, the functional approach views the language as a set of choices that can be varied according to the social purpose and situational context. Then, I turned to give a brief introduction to Halliday's (1985a; 1994) three functions of language: the ideational, the interpersonal and the textual. The three functions were simultaneously analysed. In terms of ideational function, for example, I drew teachers' attention to the noun groups in a sentence and examining how ideas are connected (clause complexing). In terms of the interpersonal function, for example, some samples were provided to show how

language use differed when interacting with others, and in terms of textual function, how language was used to construct logical and coherent text (see Table 5.11). The teachers responded very well with the activities given.

**Table 5.11 The interpersonal and the textual functions of language (Bruce 1996, Nunan 2008)**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>2. Interpersonal Function</b><br/> Language is used to enact social relationship: to interact with other people, to negotiate relationships and to express opinions and attitudes.</p> <p>Example:</p> <div> <ul style="list-style-type: none"> <li>• 'Let me have your mobile – I need to call my husband.'</li> <li>• 'I wonder if you'd mind awfully lending me your mobile – I need to call my husband.'</li> </ul> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <p><b>3. Textual Function</b><br/> Language is used to construct logical and coherent texts: how information is organised in a text and how connections are made across a text.</p> <p>Example:</p> <div> <p>There are several stages in the manufacturing of <b>tea</b>. First of all, the <b>top buds</b> and the <b>two leaves</b> immediately below are plucked from each plant. <u>In the next stage</u>, the <b>fresh leaves</b> are spread on racks or shelves and allowed to <b>wither</b> for between eighteen and twenty-four hours. <u>The next step</u> is to feed the <b>withered leaves</b> into a rolling machine for up to three hours in order to release the juices. <u>After this</u>, the <b>leaves</b> are fermented on trays before being fired in a furnace for thirty to forty minutes. <u>Then</u>, the <b>dried leaves</b> are cut, sifted, sorted and graded. <u>Subsequently</u>, the <b>tea</b> is packed in <b>tea</b> chests so that <b>it</b> can be exported. <u>After being tasted by experts</u> the <b>tea</b> is auctioned. <u>In the final stage</u>, the <b>tea</b> is put in packets and sold in shops.</p> </div> |

After that, I introduced the notion of context as a dual layer recognised in SFL and explained the relationships in these contexts of culture and of situation and language use. I explained that at the level of cultural context, the broad purpose of language use was realised by genre, and at the level of situational context, the language use was realised by register, which consisted of field, tenor and mode. Register analysis thus examined the relationship between a text and the situational context surrounding the creation of the text. To help the teachers gain a more understanding of the concept of genres and social purposes, for example, lists of schooling genres

were provided (see Table 4.11), and the teachers were asked to identify the genres that were related to the EAP classes they had taught. The complete material used in Session One of the intervention can be seen in Appendix D.

**Table 5.12 Social purposes and schooling genres (Derewianka 2003, p. 137)**

| Social purpose                                                   | Genre              | Sub-types                                                                                                                  | Example                                                                                                                      |
|------------------------------------------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| To provide information about a particular person, place or thing | Description        | Objective description<br>Literary description                                                                              | My Family<br>The Old Banyan Tree                                                                                             |
| To provide information about a class of things                   | Information report | Descriptive report<br>Taxonomic report<br>- Class/subclass<br>- Part/whole<br>Compare/contrast report<br>Historical report | The Rain Forest<br><br>Different Types of Planes<br>The Parts of a Clock<br><br>Bats and Birds<br><br>Dinosaurs              |
| To tell someone how to do something                              | Procedure          | Instructions<br>Experiment<br>Directions<br>Regulations                                                                    | Making Nasi Goreng<br><br>Changing solids to Liquids<br><br>How to Get to Hyde Park<br><br>Our Class Rules                   |
| To tell what happened                                            | Recount            | Personal recount<br>Factual recount<br>Biographical recount<br>Autobiographical recount<br>Historical recount              | My Holiday in Sydney<br>Thief Steals Car<br>The Life of Gandhi<br>My Childhood<br>The Qin Dynasty                            |
| To explain how or why a phenomenon takes place                   | Explanation        | Sequential explanation<br>Casual explanation<br>Factorial explanation<br>Consequential explanation<br>Exploration          | Life Cycle of a Frog<br>Why Hot Air Rises<br>The Causes of Erosion<br>The Effects of Poverty<br>Why Dinosaurs Became Extinct |
| To explore the human condition through storying                  | Story genres       | Narrative<br>Moral tale or fable<br>Anecdote                                                                               | The Disappearing dogs<br>The Fox and the Crow<br>When I Went to the Dentist                                                  |
| To respond to a literary text or artistic work                   | Response genre     | Personal response<br>Review<br>Interpretation                                                                              | My Favourite TV show<br>Novel of the Year<br>Crouching Tiger as a Metaphor                                                   |
| To mount an argument                                             | Exposition         | Critical response<br>Persuasion<br>Discussion/debate                                                                       | Is Rap Really Subversive?<br>War is Immoral<br>The Pros and cons of Living Alone                                             |

## Session Two

Session Two was the follow-up to genre and register. The focus of this session was on activities designed to deepen teachers' engagement with genre and register activities. These activities were done through text analysis which required students to identify: the purpose of a text, the stages required to achieve the purpose, the

situation where that kind of text likely occurred, the topic of the text (what the text was about), the relationship between people involved in the text (equal/unequal status), and mode of communication used in the text (spoken, written, face-to-face, telephone conversation and others). To attract teachers' attention to the content or of the essay, for example, activities were focused on identifying the content-carrying words such as noun and verbal groups, and also on the construction of these groups in a single clause and clause complex. In these activities teachers were asked questions such as: "How many clauses are there in paragraph one?", "How many sentences are there consisting of single clause?", "How many sentences are there consisting of two or more clause combinations?", "Look at the noun groups in paragraph two; what kind of nouns are used?" or "How many nominalisations can you find?". Similarly, to focus attention on the flow of information in a passage, activities were focused on the identification of Theme. For these activities, some instructions provided to guide the teachers were: "Underline either the first adverbial or noun groups you find on each clause and name it Theme", "Underline the rest part of the same clause and name it Rheme", "How many of the Themes are repeated using synonyms, meronyms or pronouns?" or "Can you find a clause in the passage in which the Rheme is used as the Theme in the subsequent clause?"

Texts were drawn from a range of situations, from simple face-to-face conversations to those involving more complex texts such as discussion and argument. The activity started with a simple text on face-to-face conversation between a customer and a shopkeeper (as in Table 5.13), then it gradually built up to more complex texts such as discussion and argument essays (as in Table 5.14). These activities were started

with researcher-led whole-class activity followed by group-work, pair-work and individual-work before being closed with a whole-class discussion.

**Table 5.13 Genre and register concept activity**

|                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A: Yes Please</b><br><b>B: Can I have those two?</b><br><b>A: Yes. This one's fifty five and this one's thirty five.</b><br><b>B: And have you got ..... ?</b><br><b>A: Yes. How many would you like?</b><br><b>B: I'll take four</b><br><b>A: Right. That's four dollars ninety altogether.</b><br><b>B: Here you are.</b><br><b>A: Thank you.</b><br><b>B: Thank you.</b> | <b>What's the context of the text?</b><br><br><b>What kind of a text is it? (genre)</b><br><br><b>What are the stages of the text?</b><br><br><b>What is it about? (field)</b><br><br><b>Who is involved? (tenor)</b><br><br><b>Mode of communication? (mode)</b> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**Table 5.14 Genre and register concept activity: sample of discussion essay (adapted from Bruce 1996)**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Is living in the country better than living in the city?</b></p> <p>In discussing whether living in the country is preferable to living in the city, it is necessary to look at the advantages of both. Although some people claim that the country offers a better physical environment with friendlier people, others argue that the social amenities and public transport services are superior in the city.</p> <p>There are several reasons why country-life is superior to city-life. Firstly, the physical environment in the country is cleaner, quieter and more aesthetically pleasing than that in the city. Cities lack space and greenery and are very noisy. The small number of parks and gardens can hardly compensate for the unsightly concrete buildings, noisy crowds and traffic which dominate the city landscape. In addition, the people in the country are usually friendlier than those in the city. While meeting people in cities can be difficult, in the country it is easier to establish a social life in the community. This is because a sense of community spirit exists.</p> <p>On the other hand, there are a number of advantages to living in the city. One advantage is that cities have better public transport services. Thus, the need for private transport is reduced. Public services are not as well-developed in the country as they are in the city and, consequently, it is necessary to have private a private car. Furthermore, the social amenities in the city are far superior to those in the country. Whereas theatres, restaurants, sports clubs and other entertainment facilities are usually abundant in cities, they are rather limited in the country. Therefore, it is necessary to travel to the city to take advantage of such facilities.</p> <p>In conclusion, although there are some obvious advantages to living in the country, the social facilities and public services that exist in cities are far superior to those in the country. Therefore, it can be said that living in the city is better than living in the country.</p> | <p>What is the purpose of the text?</p> <p>What genre does the text belong to?</p> <p>How is the text organised?</p> <p>What are the distinctive language features of the text?</p> <p>Field?</p> <p>Tenor?</p> <p>Mode?</p> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### **Session Three**

The third session introduced the system of Appraisal. To begin with, I gave a brief review of the concepts of context and metafunction that had been discussed in the previous sessions. Then I introduced Appraisal theory starting with ATTITUDE subsystems (Affect, Appreciation and Judgement). As research in Appraisal shows, quality academic essays are characterised by a high number of Appreciation of *valuation*. Thus, I wanted to draw teachers' attention to how the ATTITUDE subsystems, particularly Appreciation resources, were used in academic texts. After that, the session continued with GRADUATION subsystems to show how evaluation resources of Affect, Appreciation and Judgement can be graded up and down. Then, I continued with explanations of ENGAGEMENT resources to show the arguability of opinion or evaluation in writing. The practice session consisted of activities focusing on identification of Appraisal items, starting from simple sentences and building up to argument essays as seen in Table 5.15 (see Appendix E for the detailed material used in the intervention).

**Table 5.15 Exercise 5: Appraisal resources finding activities (the text is taken from Jakeman and McDowell 2003)**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Exercise 5</b></p> <p>Work individually and identify all the Appraisal resources in the text below. After that discuss your findings with your neighbor.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <p>Some people argue that the state does <i>not have the right</i> to make parents immunise their children. However, I feel the question is not whether they should immunise but whether, as members of society, they <i>have the right not to</i>.</p> <p>Preventative medicine has proved to be the most effective way of reducing the incidence of fatal childhood diseases. As a result of the widespread practice of immunising young children in our society, <i>many lives have been saved and the diseases have been reduced to almost zero</i>. In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had immunity, <i>the diseases spread easily</i>. Diseases such as dysentery were the result of poor hygiene but <i>these have long been eradicated since the arrival of good sanitation and clean water</i>. Nobody would suggest that <i>we should reverse this good practice now</i> because dysentery has been wiped out.</p> <p><i>Serious diseases such as polio and smallpox have also been eradicated</i> through national immunisation programmes. In consequence, children not immunised are far less at risk than they would otherwise be. Parents choosing not to immunise are <i>relying on the fact that the diseases have already been eradicated</i>. If the number of parents choosing not to immunise increased, there would be <i>a similar increase in the risk of the diseases returning</i>.</p> <p>Immunisation is <i>not an issue like seatbelts which affect only the individual</i>. A decision not to immunise will have <i>widespread repercussions</i> for the whole of society and for this reason, I do <i>not believe</i> that individuals <i>have the right</i> to stand aside in this disease-free society. In my opinion immunisation should be obligatory.</p> |

After the teachers had finished doing the exercise, the Appraisal resources of the texts were discussed as a whole-class activity. For example, during the discussion of Exercise 5, I asked some follow-up questions such as:

*What or who is being evaluated in paragraph two? What ATTITUDE resource is most frequently used in the evaluation? Why?*

*How is the evaluation graded up or down in paragraph two? What ATTITUDE resource is used?*

*Look at the first sentence of paragraph two. When the writer says, “Preventative medicine has proven to be the most effective way of reducing the incidence of fatal childhood diseases”: What ENGAGEMENT resource is employed? How strongly is the opinion made? Does it open for other alternative opinions?*

In this session, difficulties about the Appraisal terms that teachers had to memorise arose. While teachers succeeded in understanding and recognising the Appraisal resources in students’ texts in terms of ATTITUDE, ENGAGEMENT and GRADUATION, understanding the Appraisal sub-systems and memorising the terms used was quite challenging for the teachers, particularly because of the time allocation, which was only two hours. However, as I stayed at the language school everyday the whole four weeks, the teachers could talk about things that were unclear to them outside of the training sessions.

#### **Session Four**

The last session of the intervention program focused on the language features identified as important in successful essays activities. These activities were designed to give teachers practice in identifying the features of student essays based on the newly-designed SFL-informed rubric previously explained (see Tables 4.6-4.8 above for the marking criteria used). Informed by SFL rubric, the assessment of student essays began with broad criteria. It started with genre level to identify the overall social purpose of the writing and the stages and phases used to achieve the purpose (Nunan 2008), as described on the top layer of the rubric. For this practice, the teachers were given sample essays and were asked to identify the criteria such as those relating to the staging of the genre. The stages of a prototypical argument used

were: *Thesis ^ Arguments ^ Restatement*, and for discussion the stages identified were: *Issue ^ Sides ^ Resolution* (Coffin 2004, Martin 2009, Rose 2005, Rose and Martin 2012). Then the focus of activities proceeded to register-related features. The criteria finding session started with a whole-class activity followed by pair-work and individual work activities, before being closed by a whole-class discussion again. Similar to three previous sessions, scaffolding was strong in the beginning but it was gradually reduced as the teachers had gained more confidence (Feez 2002). The detailed activities and marking criteria can be seen in Appendix E.

## **5.4 Summary of the chapter**

The purpose of this chapter was to analyse and identify shared features of successful student texts and employ them in the design of the new marking rubric and its implementation in the intervention. The essence of the findings is that quality student essays inherit several fundamental features. Firstly, the essays have clear social purposes that are realised by stages and phases to achieve the purposes. Secondly, they are found appropriate in three main functions of language recognised in the Systemic Functional Linguistic theory: the content or field of the essays are appropriate in that the topics being discussed clearly answer the questions given; the word choices used to evaluate the issues under discussion are quite formal with less personal tone; and there is cohesiveness in the texts in that the ideas flow smoothly both between clauses as well as between paragraphs. Thirdly, the essays are free from mistakes such as spelling, punctuation and grammatical mistakes. These features were accommodated into the SFL-informed rubric under three main marking criteria: Genre, Register variables, and Accuracy. These were implemented

in the intervention, which was conducted through a four-session teacher training. The next chapter presents the results of the analysis of teachers' experiences in using the existing and new marking rubrics.

## **Chapter Six**

### **Teachers' perceptions and practices**

Thus far, the SFL theory (Halliday 1978; 1985a, Halliday and Hasan 1985, Halliday and Matthiessen 2004, Martin 1984; 1991) has informed the analysis which identified the shared characteristics of quality essays and, at the same time, shaped the design of the SFL-informed rubric and the professional learning program. This chapter presents the analysis and interpretation of teachers' accounts in relation to their marking practices. The relevant tools here are the system of Appraisal (Martin and Rose 2007, Martin and White 2005, White 2002; 2012). The chapter responds to the questions posed in Chapter One. Section One is concerned with RQ1: *How do raters perceive the existing marking criteria?* and RQ2: *What are the main sources leading to rater inconsistency in student essays?* This section describes teachers' attitudes towards the marking rubric used and the perception of consistency among their colleagues, and also identifies sources of inconsistency in marking student essays. Section Two is structured around RQ3: *How do raters' marking practices shift after using the SFL-informed criteria?* and RQ4: *How do raters respond to the SFL-informed criteria?* It considers the impact of the research intervention on teachers' practices.

#### **6.1 Teachers' opinions and practices: Pre-intervention**

This section presents the teachers' opinions of their original marking practices. In this study, Appraisal was used to analyse teachers' talk about the rubrics and the intervention because it offers a way of exploring, describing and explaining the way

language is used to evaluate, to adopt stances, and to manage interpersonal positioning and relationships (White 2002; 2012). In this case, I wanted to investigate teachers' opinions of the existing marking criteria and their marking consistency.

As discussed previously (see Chapter Four), the Appraisal taxonomy makes it possible to identify the linguistic forms realising different aspects of speakers' attitudes. ATTITUDE is seen as a resource concerned with the writer's/speaker's subjective value positions. In other words, the opinions made were closely related to the speakers' personal feelings about the marking rubric they used, their marking consistencies and marking routines. ENGAGEMENT deals with the way a given position is presented in relation to other positions. For example, when expressing feelings towards the existing marking rubric and their assessment practices, the strength of teachers' convictions can be seen as to whether they made bare assertions or statements as taken for granted (*monogloss*) or whether they recognised and made room for the existence of other opinions (*heterogloss*). The use of *heteroglossia* in teachers' accounts, for example, suggests whether the teachers were reflective, recognising that such matters are unsettled, that there is more than one way to solve the problem. GRADUATION offers a means of describing the strength of speakers' opinions, and in this research it helps identify the intensity of the teachers' perceptions regarding the issues under discussion.

As we shall see, the teachers' expressed confidence in their marking practices but saw major problems with the existing rubric. When evaluating this rubric, the teachers gave more negative opinions about it e.g., *quite difficult to define, not really*

*clear and doesn't have space to assess the content.* In contrast, when evaluating their marking practices, the teachers gave more positive opinions about them e.g., *do that based on the rubric, give a fair assessment and try my best to stick on the rules.* Based on the linguistic choices the teachers made when attributing the sources of problems in marking student essays, the majority of the negative statements were coded as Appreciation (i.e., the evaluations of things, which in this case is the marking rubric), whereas the majority of the positive statements were coded as Judgement (i.e., the evaluations of human behaviours). Affect, on the other hand, constituted only a small number of the teachers' accounts. As the foci of teachers' evaluations are an object (the rubric) and behaviour (their practices), the infrequent number of comments sourced from emotive feelings such as *like, worried* or *nervous* (Affect) is not unexpected. See Figure 6.1 for the trend of positivity and negativity in the teachers' opinions.

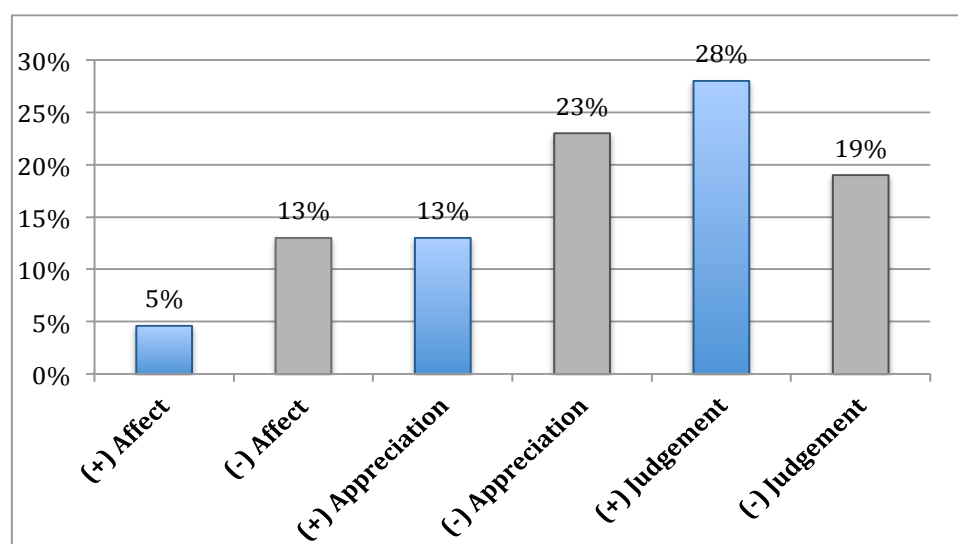


Figure 6.1 Attitudinal choice: positive and negative perceptions

### 6.1.1 Teachers' opinions of the existing marking rubric

Teachers' negative evaluation of the rubric is evident in their prevalent employment of negative ATTITUDE (Appreciation and Affects). The teachers' positive and negative perceptions of the existing marking rubric as summarised below (see Table 6.1) show that more than two-thirds (67%) of the Affect and Appreciation elements identified in the teachers' statements had negative connotations. For example, the teachers said, "It is not simply about S-V Agreement", "Grammar should be here", "They could be separated as well" and "It's not here though ... the content". In contrast, less than one-third (33%) of the teachers' comments about the rubric were in the form of positive evaluations, such as " So, I think it does help", "The rubric is very helpful" and "The rubric is very well established".

**Table 6.1 Teachers' positive-negative perceptions of the existing marking rubric**

| Affect |      |       |      | Appreciation |       |       |       |      |       | Total |
|--------|------|-------|------|--------------|-------|-------|-------|------|-------|-------|
| - sat  | +sat | -sec  | +sec | -reac        | +reac | -comp | +comp | -val | +val  |       |
| 8      | 3    | 14    | 5    | 1            | 1     | 36    | 12    | 2    | 10    | 92    |
| 8.7%   | 3.3% | 15.2% | 5.4% | 1.1%         | 1.1%  | 39.1% | 13%   | 2.2% | 10.9% | 100%  |

Overall, the teachers found that the existing marking rubric was lacking, particularly in the way it was constructed and the criteria identified as important. As indicated in Table 6.1 above, two-thirds (67%) of teachers' opinions about the rubric conveyed negative attitudes while only one-third (33%) of the attitudinal items identified in their accounts contained positive attitudes. The prevalence of negative evaluations in the teachers' accounts can be seen as their dissatisfied feelings with the existing marking rubric in terms of selections of criteria (e.g., *Subject-Verb Agreement* and the conflation of *Vocabulary* with *Spelling* and *Punctuation*, *Readability* etc. as seen

in Figure 5.6 in the previous chapter) as well as the absence of a criterion which focused on the content of student essay. The teachers' positive opinions were mainly attributed to criteria such as *General Statement* and *Thesis Statement*. Interestingly, the teachers also expressed positive views about the overall helpfulness of the rubric (e.g., *quite helpful, helps us to be on the same page and it does help*).

Teachers used Appreciation resources to evaluate the rubric; language choice considered more 'objective' than direct expressions of emotive feelings such as Affect. It is perhaps not surprising that the teachers provided a more measured and reflective response to the questions about the existing marking rubric as they evaluated it as an artefact. Their choice of Appreciation over the more direct expressions of attitudes as Affect suggests that they spent some time reflecting on the rubric and their experience as teachers in this area. Some excerpts of teachers' direct evaluation of the rubric are presented below. In Excerpts 1-3, for example, the teachers talked about the lack of space for *content* in the rubric.

|   |                                                                                                                |          |       |  |
|---|----------------------------------------------------------------------------------------------------------------|----------|-------|--|
| 1 | It <b>doesn't</b> give me a <b>not a chance</b> but <b>sort of like</b> space to assess the content.(Dina, 33) |          |       |  |
|   | deny                                                                                                           | t: -comp | focus |  |

|   |                                                                          |      |          |  |
|---|--------------------------------------------------------------------------|------|----------|--|
| 2 | The rubric <b>doesn't</b> have a space to assess the content. (Dina, 37) |      |          |  |
|   |                                                                          | deny | t: -comp |  |

|   |                                                                        |      |          |  |
|---|------------------------------------------------------------------------|------|----------|--|
| 3 | And I'll say... <b>it's not</b> here though, the content. (Jelita, 48) |      |          |  |
|   |                                                                        | deny | t: -comp |  |

Similarly, in Excerpts 4 and 5, the rubric was evaluated negatively due to the lack of space for essay *structure* in the marking criteria.

|   |                     |        |        |                 |              |
|---|---------------------|--------|--------|-----------------|--------------|
| 4 | Since the structure | is not | really | stated clearly, | (Mirzan, 24) |
|   |                     | deny   | force  | negation: +comp |              |

|   |                |          |                                                 |             |
|---|----------------|----------|-------------------------------------------------|-------------|
| 5 | we need to add | at least | one other feature, the overall structure of the | (Mirzan 25) |
|   | entertain      | force    | t: -comp                                        |             |

In Excerpts, 6-8, teachers complained about the inclusion of *subject-verb agreement* as a criterion while other grammatical features were not included in the rubric.

|   |                                         |               |                                             |             |
|---|-----------------------------------------|---------------|---------------------------------------------|-------------|
| 6 | When it goes to Subject-Verb agreement, | I don't think | this is supposed to be one of the criteria. | (Jelita, 5) |
|   |                                         | deny          | negation: +comp                             |             |

|   |           |        |                               |             |
|---|-----------|--------|-------------------------------|-------------|
| 7 | It is not | simply | about subject-verb agreement. | (Yusuf, 27) |
|   | deny      | focus  | negation:+comp                |             |

|   |                           |           |         |                                                           |               |
|---|---------------------------|-----------|---------|-----------------------------------------------------------|---------------|
| 8 | Subject-verb agreement is | important | but     | we can break it down into other parts of grammar as well. | (Whitney, 14) |
|   | pronounce                 | +val      | counter | t: -comp                                                  |               |

Lastly, in Excerpts 9 and 10, teachers complained about the conflation of *vocabulary* with *spelling* and *punctuation* into one criterion, and the lack of score traits (information on the characteristics of each score) respectively.

|   |                                      |                    |                   |             |
|---|--------------------------------------|--------------------|-------------------|-------------|
| 9 | Vocabulary, spelling and punctuation | I think they could | separated as well | (Yusuf, 28) |
|   |                                      | be                 |                   |             |
|   |                                      | entertain          | -comp             |             |

|    |       |       |         |             |
|----|-------|-------|---------|-------------|
| 10 | It is | just  | number. | (Ratna, 29) |
|    | mono  | focus | -comp   |             |

As seen in Excerpts 1-10, all the teachers' negative opinions were aimed towards the construction of the existing marking rubric and criteria used. To be more specific, the teachers expressed their concerns about the choice and arrangement of criteria, the lack of space for a content-focused criterion and the absence of information that distinguished quality or characteristic of each score, which ranged from 1 to 5, in the rubric.

The teachers' negative opinions of the compositional aspects of the rubric such as "It is not simply about subject-verb agreement", "I don't think subject-verb agreement should be one of the criteria" or "Since the structure is not really stated clearly" tended to be 'strong' in that they were formulated in a defensive way (*Disclaim* in Appraisal terms), suggesting that these are, perhaps, what really annoyed them. To elaborate, the teachers spoke of what was not in the rubric but should have been, such as "The rubric doesn't have ..." or "The structure is not stated clearly ...". Others were concessive (*Disclaim/counter* in Appraisal terms), indicating the teachers had strong opinions about these features e.g., "even though some minor aspects are not included" or "but the proportion is not clear enough".

From a dialogic perspective (ENGAGEMENT in Appraisal terms), the teachers' use of negation functions to temper their opinions because it acknowledges the alternative positive position, while simultaneously rejecting it (Martin and White 2005). This suggests teachers' knowledge about the field of academic writing and their skills to determine students' needs in terms of strengths and weaknesses (Matsuda, Saengkhum et al. 2013). For example, when one of the teachers (Jelita) said, "*Subject-verb agreement* was not part of our problems", she introduced the possibility that "*Subject-verb agreement* was part of the problems" but chose to reject such proposition. This is perhaps because of her knowledge of what students can or cannot do in academic writing classes. In other words, her statement is dialogic in that it invokes and presents itself as responding to such claim or belief that *subject-verb agreement* might have been or in other instances of her experiences part of the problem. The concessive instances in the teachers' responses also produce similar effects, i.e., the statements produce counter-expecting effects. For

example, when one of the teachers (Ratna) said, “It’s clear, even though some minor aspects are not included”, the statement functioned to counter the negative comment, implying that some other features should be included or added to the original rubric irrespective of their minor functions. Equally, this might have been a polite way to criticise the existing marking rubric and at the same time, ‘an invitation’ to discuss the particular issue.

In contrast, some teachers had stronger opinions of the rubric. The *monoglossically*-formulated statements employed by some teachers show these. In Excerpt 10, for example, Ratna’s negative opinion can be traced as her dissatisfied feeling of the scores in the original rubric, in which explanations of the score traits were not provided. Thus, when she said, “It **is** just number”, she was quite adamant about the limitations of the rubric. In brief, this statement is strong in that no room for other alternative propositions was provided, and the speakers’ bare assertion was treated as objective or ‘factual’ (Hyland 2009, p. 784).

It is interesting that one of the teachers (Mirzan) suggested the importance of adding the “overall” structure of the essay to the marking rubric. This reasonable suggestion was, however, expressed in a tentative way (*Entertain* in Appraisal terms), indicating either his uncertainty or his openness to other opinions. Mirzan’s knowledge of the importance of text-level organisation suggested a disjunction between assessment and teachers’ knowledge about language (KAL) and pedagogy (Knoch 2009). Mirzan’s comment can be seen as indicative of his awareness and recognition of the shortcomings of the rubric used in terms of its recognition of the stages necessary to achieve the goal of essay writing, which, in this particular

context, refers to ‘arguing on the positive and negative impacts of globalisation on developing countries’. To elaborate, Mirzan seemed to realise that if the EAP students were asked to argue and convince the readers (e.g., the teachers who marked the essays) on the topic given (e.g., positive-negative impacts of globalisation on developing countries), their work (essays), should have been assessed based on the schematic structure used to achieve the purpose of the essay. This is an issue that was discussed in the design of the intervention, in which a criterion focusing on the overall structure of the essay was included in the SFL-informed rubric under the *Genre: social purpose and stages developed in essays* (see Chapter Five).

Interestingly, despite the teachers’ dissatisfaction with aspects of the existing marking rubric, they declared the rubric to be useful in their marking, combining positive lexis with unequivocal declaration of its value. This may suggest the teachers’ loyalty to the institution and at the same time, their reluctance to challenge the policy or the accepted practice. The most common evaluative words used were *helpful* or *help*. In other words, the teachers evaluated the helpfulness of the rubric to assist them to be ‘on the same page’ and to be less subjective in their marking. Drawing on GRADUATION resources (the degree of values) of *force/intensification*, most of the teachers commented positively, even enthusiastically on the helpfulness of the rubric. Similarly, when considering the rubric from a dialogic perspective (ENGAGEMENT in Appraisal terms), the teachers’ positive valuation was enforced by the prevalent use of single-voiced locutions (*monogloss*) in their accounts. They were unequivocal in their support, e.g., “It **is** very helpful” and “the rubric **is** very

helpful”. These opinions were emphasised in the frequent employment of the pre-modified adverb *very* (*force/intensification* in Appraisal term).

|    |               |                                 |             |               |
|----|---------------|---------------------------------|-------------|---------------|
| 11 | It is         | very                            | helpful,    | (Mirzan, 10)  |
|    | mono          | force                           | +val        |               |
| 12 | the rubric is | very                            | helpful     | (Mirzan, 11)  |
|    | mono          | force                           | +val        |               |
| 13 | So I think    | it does                         | helps.      | (Dina, 57)    |
|    | entertain     | force                           | +val        |               |
| 14 | It            | helps us to be on the same page | so to speak | (Whitney, 26) |
|    |               | force/t: +val/pronounce         |             |               |

On the other hand, when asked if the existing marking rubric fostered consistency, one teacher responded as below:

|    |                    |       |          |             |
|----|--------------------|-------|----------|-------------|
| 15 | Yeah, yeah they're | quite | helpful. | (Yusuf, 25) |
|    | concur             | force | +val     |             |

This indicates that the teacher agrees but with somewhat less intensity in the message carried. Dialogically, the speaker showed his solidarity by positing agreement with projected dialogic partners (White 2012), suggesting that others (his colleagues and/or I myself) shared similar feelings regarding the helpfulness of the rubric.

In addition to using direct evaluations through choices from the system of Appreciation, there is, however, a small number of the teachers' negative opinions of the rubric that were expressed through emotive feelings in realisation of Affect. Given the spoken and the informal nature of the discourse, the teachers' use of such emotional responses is not unexpected. In fact, it might be suggested that the less formal the discourse is, the more it relies on emotional responses in the evaluations

made. As seen in Excerpts 14-16 below, for example, the teachers did not say directly that some of the criteria in the rubric were ‘difficult’ or ‘not clear’. Instead, they chose to make more indirect evaluations of the rubric by focusing on the speakers’ emotive feelings such as ‘I feel, I find, and I’m’. Again, the intensity of such opinions was realised through the prevalent use of *force/intensification*. This means that despite the indirectness of their evaluations, the teachers’ opinions were definite in that they chose not to downscale the negative perceptions. In other words, the teachers used utterances such as difficult – *most* difficult, clear – *not really* clear rather than downscaling the negative attitudes (e.g., *kind of* difficult or *sort of* clear).

|    |                                                                                                                                             |
|----|---------------------------------------------------------------------------------------------------------------------------------------------|
| 16 | OK, the <b>most</b> <b>difficult</b> <b>thing</b> a couple of times, <b>is</b> the vocabulary that the students use in the essay. (Dina, 1) |
|    | force -sec force mono                                                                                                                       |
| 17 | <b>Sometimes</b> I find it <b>difficult</b> to give the score. (Yusuf, 3)                                                                   |
|    | force -sec/force                                                                                                                            |
| 18 | I <b>think</b> Readability also, I’m not <b>really</b> <b>clear</b> about Readability. (Dina, 9)                                            |
|    | entertain force neg +sec                                                                                                                    |

Further, the teachers’ choice of attitudinal words sourced from their emotion (Affect in Appraisal terms) in the interview data shows two important findings. Firstly, speakers expressed some anxiety (*in/security*) over feelings of frustration (*dissatisfaction*) e.g., *just feel something is wrong* and contentment (*satisfaction*) with the original criteria e.g., *my favourite part* (The detailed information on the Affect elements used has been summarised earlier in Table 6.1). This finding suggests that the teachers’ perceptions of the marking rubric were closely related to their lack of confidence of using the rubric. Secondly, in terms of the two sub-systems of Affect identified, the total number of negative feelings employed was overtly higher than the positive ones. The high occurrence of negative emotional

reaction (Affect) with respect to the construct of the rubric suggests the teachers' timidity and concerns of their marking quality using the existing rubric. The teachers' use of explicit emotive feelings (e.g., *difficult* and *not clear*) as in excerpts 16-18 above, do reveal that the teachers expressed their anxiety openly.

In summary, the teachers drew on the resources of Appreciation to describe the existing rubric, identifying features such as *General Statement* and *Thesis Statement* as important and positive features of the rubric. They pointed out its shortcomings such as the use of a limited grammatical feature (*s-v* agreement) as a criterion, the conflation of *vocabulary* with *spelling* and *punctuation*, and the unavailability of a criterion that focused on the content of student essays. Interestingly, despite revealing their dissatisfaction with these aspects, the teachers rated the rubric overall as useful.

### **6.1.2 Teachers' opinions about their consistencies and practices**

Teachers tend to be less consistent in marking students' vocabulary in terms of range and accuracy (Ruegg, Fritz et al. 2011) and, as the literature has shown, teachers' judgements of their capabilities can be strongly reflected in their assessment practices (Hyland 2013). As discussed in the previous chapter, the resources of Judgement serve to evaluate human behaviour by reference to socially constructed norms and regulations (Martin and Rose 2007, Martin and White 2005). In this study, judgement was relevant to the teachers' perceptions of their own efficacy; in other words, their marking capabilities and marking practices. The teachers' opinions were analysed in terms of how they perceived their capabilities in marking student essays, and how they perceived the way they mark student essays.

Table 6.2 reveals that negative *capacity* of marking consistently such as “It is hard for me to be consistent”, “I find it hard in this area of vocabulary” or “I think ... especially vocabulary” and positive *propriety* with respect to assessment practices such as “I do mark according to these criteria” or “I have to do that based on the rubric” were two of the most pervasively employed Judgement resources. In other words, teachers worried about their essay marking capabilities but were compliant with the regulations set by the institution in which they worked. In contrast, marking practices were not evaluated in terms of *veracity* or truthfulness, suggesting that this is assumed. The pervasive use of these two resources suggests that the teachers’ marking practices were perceived and developed in terms of their capability and ethical qualities. The majority of teachers’ statements regarding their practice in these terms were expressed in dialogic formula (*heterogloss*) while only a few were expressed unequivocally (*monogloss*). This suggested the tentativeness of the opinions made, that it might be different from one to another and thus, make more room for negotiation.

**Table 6.2 Teachers' use of Judgement in terms of capabilities and practices**

| Judgement |       |       |       |      |      |      |      |       |       | Total |
|-----------|-------|-------|-------|------|------|------|------|-------|-------|-------|
| -norm     | +norm | -cap  | +cap  | -ten | +ten | -ver | +ver | -prop | +prop |       |
| -         | 2     | 31    | 14    | -    | 2    | -    | 6    | 1     | 25    | 81    |
| -         | 2.5%  | 38.3% | 17.3% | -    | 2.5% | -    | 7.4% | 1.2%  | 30.1% | 100%  |

As the data above shows, while the teachers had a strong sense of propriety in terms of their institutional practices, they had misgivings about their capacity to mark students’ essays consistently. As seen in Table 6.2, the number of positive Judgements of *propriety* was substantially high (30.1%), but low in *capacity* (17.3%). This may suggest the teachers’ endeavour to be virtuous in their practices.

The high use of positive *propriety*, as seen in Excerpts 19-23, for example, functioned as affirmations of teachers' convictions of the way they usually mark student essays. In the excerpts below, for example, the teachers' use of the phrases such as *according to*, *stick to the rules* and *based on* functions to indicate their abidance with the institutionalised rules.

|    |                       |                                                     |              |
|----|-----------------------|-----------------------------------------------------|--------------|
| 19 | I DO mark             | according to these criteria.                        | (Whitney, 1) |
|    | pronounce             | t: +prop                                            |              |
| 20 | I try my best         | to stick to the rules... the list.                  | (Yusuf, 1)   |
|    | force                 | +prop                                               |              |
|    | pronounce             |                                                     |              |
| 21 | I check               | everything.                                         | (Jelita, 17) |
|    | mono                  | +prop                                               |              |
| 22 | I mark the essays     | according to the question; what the essay requires. | (Ratna, 1)   |
|    | mono                  | t: +prop                                            |              |
| 23 | and I have to do that | based on the rubric.                                | (Dina, 35)   |
|    | entertain             | t: +prop                                            |              |

Their preference for more institutionalised Judgement indicates the acute awareness of high ethical stance expected of them as assessors of student essays. In other words, the teachers' evaluations of their ethics in the marking of student essays were very positive despite their perceived negative opinions of some of the criteria in the rubric used. As discussed earlier (see Section 6.1.1 above), teachers complained about the limited grammatical feature, the lack of a content-focused criterion, the combination of *vocabulary* with *spelling* and *punctuation* and the lack of clarity of some criteria use. Thus, the prevalence of positive *propriety* in teachers' comments may suggest that they were reluctant to challenge the institutional marking policy, and instead, chose to conduct their marking practices according to directives made by an institution (the language school they worked in).

The teachers' endeavours to be ethical were further informed by the prevalent use of single-voice expressions (*monogloss*) that reveal much about their industry: "I **check** everything", "I **mark** the essay according to". These declarations affirm the truthfulness and commitment to teachers' work (Martin and White 2005). The teachers' endeavours to be ethical were not only shown by their unequivocal declarations but also by their coupling them with action processes. This described their labour in assessing. In addition, the teachers' perceived positive ethics were realised in the prevalent employment of 'high degree' adverbs functioning to show the certainty (*heterogloss/proclaim/pronounce*) such as "I *DO* mark" and also high degree of modality to show obligation (*heterogloss/entertain*) such as "I *have to* do that based on the rubric". Whitney's use of a finite auxiliary 'do' in Excerpt 19, for example, served to add emphasis to the authorial voice (Hyland 2009, Martin and White 2005) and, at the same time, realised her value position.

In contrast to the positive perceptions of their actual marking practices, most of the teachers perceived their marking capabilities negatively. This is shown by the high employment of negative *capacity* with respect to marking students' essays consistently. As seen in Table 6.2 above, negative *capacity* was chosen frequently by the teachers, in that the number was double that of the positive ones. Although positive *capacity* was found in their statements (see Excerpts 24-25), the instances of negative Judgement of *capacity* identified, such as in Excerpts 26-29, were prevalent, pointing to the teachers' awareness of problems in marking students' essays in relation to their dissatisfaction with the selection of criteria as discussed earlier (e.g., the use of *S-V Agreement* as a criterion and the conflation of *Vocabulary* with *Spelling* and *Punctuation*).

|    |                         |                                                     |  |  |  |  |            |
|----|-------------------------|-----------------------------------------------------|--|--|--|--|------------|
| 24 | I think I am consistent | with my marking, with the way I look at the essays. |  |  |  |  | (Ratna, 9) |
|    | entertain               | +cap                                                |  |  |  |  |            |

|    |         |            |           |        |                             |              |
|----|---------|------------|-----------|--------|-----------------------------|--------------|
| 25 | but     | in general | I find it | pretty | close to the other teacher. | (Whitney, 7) |
|    | counter |            |           | force  | t: +cap                     |              |

|    |                   |                            |  |  |  |            |
|----|-------------------|----------------------------|--|--|--|------------|
| 26 | it is hard for me | to be consistent you know. |  |  |  | (Dina, 12) |
|    | pronounce         | neg: +cap                  |  |  |  |            |

|    |           |                                   |  |  |  |              |
|----|-----------|-----------------------------------|--|--|--|--------------|
| 27 | I find it | hard in this area for vocabulary. |  |  |  | (Jelita, 36) |
|    | pronounce | -cap                              |  |  |  |              |

|    |           |           |           |                    |  |  |            |
|----|-----------|-----------|-----------|--------------------|--|--|------------|
| 28 | Sometimes | I find it | difficult | to give the score. |  |  | (Yusuf, 3) |
|    | force     | pronounce | -cap      |                    |  |  |            |

|    |         |           |          |        |             |          |             |             |
|----|---------|-----------|----------|--------|-------------|----------|-------------|-------------|
| 29 | But     | sometimes | I myself | cannot | avoid being | a little | subjective. | (Mirzan, 2) |
|    | counter | force     |          | deny   |             | force    | -cap        |             |

Overall, the teachers' negative evaluations of their capabilities in marking students' essays consistently were predominantly explicit, as shown by the employment of Judgement sub-systems in their talk about perceived consistencies in marking. For example, the six teachers' accounts of the marking consistencies were made predominantly on the basis of capability (Judgement/*capacity* in Appraisal terms) directed to themselves as the essay assessors (e.g., *I think I am consistent*, *very difficult to decide*, *I might not be consistent* and *I find it difficult*). This suggests teachers do not shirk from taking responsibility for their marking practices. In Excerpts 24-29 above, for example, the directness of evaluations of human, and by definition, their behaviours was signalled by their focus on using 'I' or 'me'. As far as the positioning of the teachers' voices is concerned (ENGAGEMENT in Appraisal terms) it is notable that quite a few instances were found, such as *pronounce*. The ubiquities of *pronounce* elements such as "I find it hard"; "I find it difficult" and "It is hard for me" in the teachers' accounts indicate their maximal investment or authorial voices in the propositions made (White 2012).

Reflections on the existing rubric and their marking practices, inevitably led the teachers to reflect on students' writing skills. Some of the examples are seen below:

|    |                                                                              |       |      |       |           |  |
|----|------------------------------------------------------------------------------|-------|------|-------|-----------|--|
| 30 | My students sometimes are not really sure about Thesis Statement. (Dina, 45) |       |      |       |           |  |
|    |                                                                              | force | deny | force | neg: +cap |  |

|    |                                                       |       |           |  |  |  |
|----|-------------------------------------------------------|-------|-----------|--|--|--|
| 31 | They don't really understand the issue. (Whitney, 29) |       |           |  |  |  |
|    | deny                                                  | force | neg: +cap |  |  |  |

|    |                                                           |       |             |  |  |  |
|----|-----------------------------------------------------------|-------|-------------|--|--|--|
| 32 | They are not really thinking for themselves. Whitney, 32) |       |             |  |  |  |
|    | deny                                                      | force | (neg: +cap) |  |  |  |

As seen in the excerpts above, all of the evaluations were targeted at students and their capabilities such as “My students are not really sure about *Thesis Statement*”, “They don’t really understand the issue” or “They are not really thinking for themselves”. Judging from what the students were claimed capable of or not, there seems to be a strong correlation between teachers’ perceptions of students’ problems, teachers’ values as the essay markers and the specific criteria on the marking rubric. For example, in Excerpts 31 and 32 Whitney’s perception of her students’ lack of capability in understanding the issue in independent thought seems to stem from her emphasis on the importance of the Thesis Statement. Whitney holds that argument essay writing is very ‘thesis driven’ in that the *Thesis Statement* dictates the content of the essay or how the issue will be approached (Rose and Martin 2012). In order to be able to provide the preview that reflects the whole content of the essay, students, as Whitney proposed, need to have good understanding of the issue under discussion (e.g., positive and negative effects of globalisation on developing countries). Her teaching practice also reflected this importance. Thus, her comments on students’ capabilities suggested a strong link between teaching and assessing practices. This is reflected below:

|    |           |                      |                    |            |               |
|----|-----------|----------------------|--------------------|------------|---------------|
| 33 | Probably  | the Thesis Statement | the most important | part of an | (Whitney, 20) |
|    | entertain |                      | force              | +val       |               |

Similarly, Jelita's opinions on students' capability of mastering *Subject-Verb Agreement* (see Excerpt 34) seem to be rooted in her disapproval of the inclusion of *S-V Agreement* and her endorsement of changing this criterion into a more general grammar or sentence structure (see Excerpt 35). These opinions, at the same time, confirm those of Matsuda, Saenkum and Accardi (2013) – that is, teachers are aware of students' needs, their proficiencies and the importance of criteria in recording this information.

|    |              |        |               |                              |             |
|----|--------------|--------|---------------|------------------------------|-------------|
| 34 | Most of them | do not | have problems | with Subject-Verb Agreement. | (Jelita, 7) |
|    | force        | deny   | -cap          |                              |             |

|    |          |         |           |            |                        |             |
|----|----------|---------|-----------|------------|------------------------|-------------|
| 35 | The only | problem | that they | might have | will be the structure. | (Jelita, 8) |
|    | force    | -cap    | entertain |            |                        |             |

Jelita found that *S-V Agreement* is not relevant to an assessing rubric because it is not problematic for students. The students were found to have a range of problems constructing their sentences including the use of the right tenses, active-passive voices, and conjunctions. This, at the same time, echoes what another teacher, Yusuf, said earlier in Excerpt 7, “So it is **not simply** about *Subject-Verb Agreement*”. Seen from a dialogic perspective, Yusuf's statement using an element of *focus* functions to upscale or sharpen his opinion, in which the softened element *simply* was negated to produce a counter-expectancy effect in its meaning. This signified his confidence in his judgement that marking grammar was *not simply* about *Subject-Verb Agreement*, and at the same time positioned himself as having greater expertise in this area than that of the person who developed the marking criteria. Taking Martin and White's (2005) perspective, this locution can then be

seen as corrections to misconceptions on the part of the addressee who was thought to believe that marking grammar **is** simply *about Subject-Verb Agreement*.

In summary, teachers' opinions of their marking practices can be grouped into two main themes: negative opinions of their capabilities and positive opinions of their ethics. That is, they lacked confidence in their skills but believed they meant well. When making evaluations of their essay marking skills, most teachers drew on Judgement resources of negative *capacity* e.g., "It's hard for me to be consistent" or "I find it difficult to give scores", suggesting their lack of capability. Despite that, the prevalence of positive Judgements of propriety, such as "I do mark according to these criteria" or "I try my best to stick to the rules" indicates teachers' endeavours to remain ethical in their marking practices. In addition, teachers' comments e.g., "My students, sometimes, are not really sure about Thesis Statement" or "Most of them do not have problems with subject-verb agreement", have shown the inseparability of assessment and students' needs.

### **6.1.3 Pre-training inter-rater reliability and sources of inconsistency**

#### **6.1.3.1 Consensus estimate of inter-rater reliability**

Analyses of teachers' perceptions and teachers' marking consistencies reveal a strong correlation between the two components, in that the teachers' negative perceptions of their consistency is reflected in their lack of marking consistency. As shown above, the teachers regularly expressed negative opinions about their capabilities of marking consistently. This was shown by dominance of negative elements of *capacity* over the positive ones e.g., "Sometimes it's kind of dilemma to

give what kind of score” or “I cannot avoid being a little subjective” or “That can be difficult” (see Table 6.2). These opinions are confirmed by the results of consensus estimates of inter-reliability counts, which were low. As seen in Table 6.3 below, the average consensus estimate was only 25.6%. This means that of all the 10 essays marked by a teacher, only a quarter of them (two and a half essays) shared similar scores with those of other teachers (see Chapter Four, Section 4.5.1 for the detailed information of consensus estimates of inter-rater reliability counts).

The data summarised in Table 6.3 shows the consensus estimates of inter-rater reliability among the six EAP teachers. Moving from left- to right-hand column, the data include names of raters, the number of essays that received similar scores among the raters, and the percentages of the similar-scored markings. For example, Dina had 36% consensus estimate of inter-rater reliability rate i.e., she had 18 out of 50 markings that shared similar scores with other teachers. Jelita, on the other hand, had only 14% – that is, only 7 out of 50 of her essay markings that received scores similar to those of her colleagues.

**Table 6.3 Pre-training consensus estimates of inter-rater reliability**

| <b>Raters</b> | <b>Similar Score Essay</b> | <b>Percentage</b> |
|---------------|----------------------------|-------------------|
| Dina          | 18/50                      | 36%               |
| Jelita        | 7/50                       | 14%               |
| Mirzan        | 12/50                      | 24%               |
| Ratna         | 15/50                      | 30%               |
| Yusuf         | 9/50                       | 18%               |
| Whitney       | 16/50                      | 32%               |
| <b>Mean</b>   | <b>77/300</b>              | <b>25.6%</b>      |

As has been argued previously in Chapter Two, a high inter-rater reliability in marking using holistic scales does not always correspond to a high agreement

among assessors. In the case of using holistic scales, for example, it may be argued that the same scores assigned to one paper or essay might come from different scores given to the sub-skills in writing such as *Grammar*, *Vocabulary*, *Content*, *Organization* and others (J. D. Brown 1991, Shi 2001). Despite the similarity of scores assigned by the teachers, they may, however, “arrive at those scores from somewhat different perspectives” (J. D. Brown 1991, p. 601). Similarly, Shi’s (2001) study shows that high inter-rater reliability does not necessarily mean that markers agree with what constructs a good essay.

In an attempt to provide a more detailed analysis of the variability of scores assigned by the six teachers, all of the ten essays were analysed in terms of the consistency rate they received. As seen below in Table 6.4, Essay 1 and Essay 7 were found to have the most agreed-upon scores among the six assessors.

**Table 6.4 Rate of recurrence of the ten essays**

| Essay    | Rate of recurrence | Percentage |
|----------|--------------------|------------|
| Essay 1  | 20 times           | 67%        |
| Essay 2  | 6 times            | 20%        |
| Essay 3  | 2 times            | 6.7%       |
| Essay 4  | 4 times            | 13%        |
| Essay 5  | 4 times            | 13%        |
| Essay 6  | 8 times            | 27%        |
| Essay 7  | 16 times           | 53%        |
| Essay 8  | 6 times            | 20%        |
| Essay 9  | 6 times            | 20%        |
| Essay 10 | 6 times            | 20%        |

In terms of frequency of scores, Essay 1 had a 67% consistency rate; it received 20 similar scores out of a maximum of 30 recurrences. Essay 7 had a slightly lower consistency rate (53%), in that it received 16 similar scores out of a maximum of 30 recurrences. As the essays with the highest inter-assessor agreement, these two essays were analysed to identify features of the essay with the most and least

agreement so as to answer the RQ2, which looks for sources in student texts which led to rater inconsistency. This will shed light on whether or not the teachers' perceived difficulties identified in the previous chapter (see Chapter Five, Figure 5.6) were correlated with their score consistencies in terms of genre stages particularly on *Topic-Supporting Sentences* and other features such as *Vocabulary* and *S-V Agreement*.

### 6.1.3.2 Score reliabilities among teachers

There is a strong correlation between what teachers perceived as problematic features in essay marking and the results of inter-rater reliability counts. In terms of genre stages, analysis of teachers' perceptions has revealed the teachers' perceived difficulties in marking *Topic Sentence* and *Supporting Sentences* as separate criteria. These perceptions were reflected by the results of teachers' consistency counts in assessing the related criteria. As seen in Table 6.5 below, the **Argument** stage or the *body* paragraphs, which consists of **topic** (identified as *Topic Sentence* in the existing marking rubric) and **elaboration** (identified as *Supporting Sentence* in the existing marking rubric) had the biggest score discrepancy in the two essays. The highest score assigned to the *Topic Sentence* in each of the two essays was 3 (three) while the lowest was 1 (one). In other words, teachers could not reach consensus in their marking of this feature of students' essays. In terms of the *Supporting Sentence*, the disparity of scores, particularly of Essay 7, was even greater. The highest assigned score was 4 (four) and the lowest was 1 (one); indicating teachers' inability to reach consensus in marking this part of students' essays. Similarly, teachers' marking of this stage of Essay 1 was found inconsistent because the

discrepancy between the highest and lowest scores was more than 1 (one) point apart – that is, the highest score was 3 (three) and the lowest was 1(one).

**Table 6.5 Score gap in terms of genre stages**

|               | Introduction      |         |                  |         | Body           |         |                     |         | Conclusion |   |
|---------------|-------------------|---------|------------------|---------|----------------|---------|---------------------|---------|------------|---|
|               | General Statement |         | Thesis Statement |         | Topic Sentence |         | Supporting Sentence |         |            |   |
|               | Essay 1           | Essay 7 | Essay 1          | Essay 7 | Essay 1        | Essay 7 | Essay 1             | Essay 7 |            |   |
| Lowest score  | 2.5               | 3       | 2.5              | 3       | 1              | 1       | 1                   | 1       | 2          | 2 |
| Highest score | 4                 | 4       | 4                | 4       | 3              | 3       | 3                   | 4       | 4          | 4 |
| Score gap     | 1.5               | 1       | 1.5              | 1       | 2              | 2       | 2                   | 3       | 2          | 2 |

As the data above shows, the level of consistency among teachers was quite low. As we have seen earlier (see Section 4.5.1), assessors are considered to have reached consensus so long as the ratings do not differ by more than one point above or below those of the other assessor (Stemler 2004). Each criterion in the rubric can be scored only from 1 to 5, on average. Some teachers gave these two phases (i.e., *Topic Sentence* and *Supporting Sentence*) the lowest possible score in the existing rubric (1 out of 5) whereas some other teachers graded them with above-average scores (3.5 out of 5). With reference to Stemler (2004) on the maximum score difference allowed for a small range of scoring system, this finding can be taken as an indication of disparity of scores among the teachers in marking this stage of genre, and at the same time affirms teachers’ perceived problems in marking *Topic-Supporting Sentences* discussed earlier.

However, the teachers’ perceptions and practices did not always go hand in hand. The teachers also had problems with marking the concluding stage of the essay consistently, e.g., the highest score was 4 (four) and the lowest was 2 (two). However, this problem was not identified in the analyses of teachers’ accounts

summarised earlier in Figure 5.6 (see Chapter Five, Section 5.5) as none of them mentioned the concluding stage of the essay as being problematic to assess. Based on the variability of scores assigned, however, the difference between the highest and lowest score the teachers assigned revealed the disparity of opinions regarding the quality of the Conclusion part of the given essays. In fact, the difference between the highest and lowest score was identical with that of the *Topic Sentences*.

Of all the problems identified in the teachers' marking of genre stages of students' essays, the teachers appeared to have the least problem in assessing the *Introduction* part of the essays. As revealed in Table 6.5 above, the teachers were able to reach consensus in their ratings of the *General Statement* and *Thesis Statement* of Essay 7. In their ratings of Essay 1, however, the variability of scores was slightly higher but relatively smaller than those of *Topic-Supporting Sentence* and *Conclusion*. This can be considered as an indication that there were fewer problems in consistency in marking the *Introduction* or **Thesis** stage of students' essays.

Despite that, the score discrepancy found with respect to the three argument stages indicates there are inconsistencies. In other words, although Essay 1 and Essay 7 received quite a high rate of agreed scores, which were 20 and 16 times respectively (see Table 6.4), overall, there is a lack of inter-rater consistency in terms of the scores assigned to the marking of genre stages. This means that despite the similarity of scores the six assessors assigned to Essay 1, they seemed to arrive at the scores from somewhat different perspectives. This finding affirms Brown (1991) and Shi (2001) who argue that the similar score assigned does not always go hand in hand with the perceived importance.

**Table 6.6 Score gap: accuracy features of essays 1 and 7**

| Accuracy features | S-V Agreement |         | Vocabulary, Spelling and Punctuation |         | Cohesion, Readability and Relevance |         |
|-------------------|---------------|---------|--------------------------------------|---------|-------------------------------------|---------|
|                   | Essay 1       | Essay 7 | Essay 1                              | Essay 7 | Essay 1                             | Essay 7 |
| Lowest score      | 3             | 3       | 2                                    | 2       | 1                                   | 1       |
| Highest score     | 4             | 4       | 3.5                                  | 3.5     | 2.5                                 | 3       |
| <b>Score gap</b>  | 1             | 1       | 1.5                                  | 1.5     | 1.5                                 | 2       |

Teachers' perceptions of assessing *Subject-Verb Agreement* were compatible with their marking consistencies in the given criterion. The analysis of teachers' accounts revealed that *Subject-Verb Agreement* was not perceived as being problematic to assess, and these perceptions were confirmed by the results of inter-rater reliability counts presented in Table 6.6. With a score gap of 1 (one), *Subject-Verb Agreement* appeared to be the feature most agreed-upon among the teachers. In the two essays analysed, the highest score assigned to this feature of students' writing was 4 (four) while the lowest was 3 (three). Based on Stemler's (2004) formula explicated above, the teachers evidently had inter-rater reliability in the marking of *Subject-Verb Agreement*, and they did not appear to have problems in marking students' essays based on this criterion. Dina, one of the teachers participating in this study, affirmed this proposition. In Excerpt 36, she appraised her capability of marking the given criterion very positively using the adjective *clear*. Two other participants, Yusuf and Whitney, also reverberated Dina's perception, and suggested the need to have a more general *Grammar* as a criterion so that other grammatical features of student writing could also be assessed (see Excerpts 37-40).

|    |                                                                                                                      |           |      |  |           |
|----|----------------------------------------------------------------------------------------------------------------------|-----------|------|--|-----------|
| 36 | To decide Subject-Verb Agreement <span style="color: blue;">would be</span> <span style="color: blue;">clear.</span> |           |      |  | (Dina, 5) |
|    |                                                                                                                      | entertain | +cap |  |           |

|    |         |                                             |                                          |                       |             |
|----|---------|---------------------------------------------|------------------------------------------|-----------------------|-------------|
| 37 | Grammar | <span style="color: blue;">should be</span> | <span style="color: green;">here,</span> | grammar or structure, | (Yusuf, 26) |
|    |         | entertain                                   | t: -comp                                 |                       |             |

|    |                                                                        |             |          |  |
|----|------------------------------------------------------------------------|-------------|----------|--|
| 38 | so it is <b>not</b> <b>simply</b> <b>about Subject-Verb Agreement.</b> | (Yusuf, 27) |          |  |
|    | deny                                                                   | focus       | t: -comp |  |

|    |                                                                         |               |      |  |
|----|-------------------------------------------------------------------------|---------------|------|--|
| 39 | <b>Yeah</b> general Grammar, Subject-Verb Agreement <b>is important</b> | (Whitney, 14) |      |  |
|    | concur                                                                  |               | +val |  |

|    |                                                                             |               |  |  |
|----|-----------------------------------------------------------------------------|---------------|--|--|
| 40 | <b>but</b> <b>we can break it down into other parts of grammar as well.</b> | (Whitney, 15) |  |  |
|    | counter                                                                     | t: -comp      |  |  |

In a similar vein, Jelita confirmed the proposition and denied that her EAP students had problems with *Subject-Verb Agreement*. She further suggested that this criterion be removed and replaced with Structure for a more generally applicable grammar.

|    |                                                                                                                 |             |          |  |
|----|-----------------------------------------------------------------------------------------------------------------|-------------|----------|--|
| 41 | When it goes to Subject-Verb Agreement, <b>I don't think</b> <b>this is supposed to be one of the criteria.</b> | (Jelita, 5) |          |  |
|    |                                                                                                                 | deny        | t: -comp |  |

|    |                                                                                    |             |      |  |
|----|------------------------------------------------------------------------------------|-------------|------|--|
| 42 | <b>most of them</b> <b>do not</b> <b>have problems</b> with Subject-Verb Agreement | (Jelita, 7) |      |  |
|    | force                                                                              | deny        | +cap |  |

|    |                                                                                        |             |           |  |
|----|----------------------------------------------------------------------------------------|-------------|-----------|--|
| 43 | The <b>only</b> <b>problem that they might have</b> <b>will be</b> sentence structure. | (Jelita, 8) |           |  |
|    | force                                                                                  | -cap        | entertain |  |

Analysis of *Subject-Verb Agreement* counts, as seen in Table 6.7 below, confirms Jelita's denial of such a problem in students' essays. In fact, of all the eighteen *Subject-Verb Agreement* elements identified in Essay 1, for example, there were only two related mistakes found, e.g., clauses no. 3 and 18. In Clause 3: *It can make [^have] a big influence to [^on] developing countries*, the *Subject-Verb Agreement* mistake was not purely related to grammatical errors (form-based grammar) such as singular/plural or past/present/future verbs. Instead, the mistake made was more closely related to meaning-based grammar i.e., the verb and preposition used resulted from the influence of the student's first language (Indonesian), in which *make* and *give an influence to* are grammatically and semantically correct. In systemic functional grammar the congruence between Subject and the verb used can

be diagnosed or probed by using a *mood tag* (see, for example Bloor and Bloor 2004, Halliday 1985a; 1994), which is also known as a *question tag* in traditional grammar. Therefore, when analysing on the basis of Subject-Verb Agreement alone the expressions *It can make*; *It can give* and *It can have* are all grammatically correct because they are all congruent with the *mood tags/question tags* used (*can't it?*).

**Table 6.7 Subject-Verb Agreement in Essay 1**

| Phase         | Subject-Verb Agreement                                                                                                                 |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------|
| (background)  | 1. In this era, globalisation has taken<br>2. It can be seen                                                                           |
| position      | 3. It can {make} {^have}<br>4. Besides that, it also gives                                                                             |
| preview       | _____                                                                                                                                  |
| topic         | 5. Globalisation can have<br>6. It shows                                                                                               |
| elaboration 1 | 7. For example, in globalisation era, education is<br>8. People need to spend<br>9. It shows<br>10. Without being realized, it becomes |
| elaboration 2 | 11. Not only that, it can also influence<br>12. It can be seen                                                                         |
| topic         | 13. Globalisation also makes                                                                                                           |
| elaboration 1 | 14. It may give                                                                                                                        |
| elaboration 2 | 15. It may also build                                                                                                                  |
| review        | 16. In conclusion, globalisation has been widespread                                                                                   |
| restate       | 17. Government needs to take part<br>18. and {make} {^play an active} role                                                             |

Similar to the Topic-Supporting Sentence, teachers' perceptions and practices in terms of marking *Vocabulary* and *Readability* were strongly correlated. The analysis of teachers' accounts has revealed that *Vocabulary* and *Readability* were perceived as two of the most problematic features of students' essays to mark along with *Topic-Supporting Sentence* and *Content* (see Chapter Five). The results of consensus estimates of inter-rater reliability counts on the given criteria confirm these perceptions. With a score gap of 1.5 each, two other features used in the existing marking rubric such as (1) *Cohesion, Readability and Relevance* and (2)

*Vocabulary, Spelling and Punctuation* were found to have no inter-rater reliability score (see Table 6.6 above) because the score discrepancy was more than 1 (one) point above or below (Stemler 2004). The highest score assigned for *Cohesion, Readability and Relevance* was 2.5 while the lowest was 1. Looking at the score range, it can be concluded that none of the teachers felt satisfied with the quality of Essay 1 on the related feature; from scale 1 to 5, the highest score given was only 2.5. However, it was yet unclear which of the three combined elements contributed to the low scores assigned to this criterion. In terms of *Readability*, for example, the teachers appeared to have their own interpretations of the feature. The literature (see, for example, Eckes 2008; 2012, Hyland and Anan 2006) has shown that the different interpretations of criteria and their importance in a rubric would inevitably lead to the different scores assigned, and eventually resulted in lack of inter-assessor consistency. Dina's statements on *Readability* (see Excerpts 47-48) for example, show how the unclear definition of such a feature had resulted in her inconsistency in marking:

|    |              |                                                     |                     |            |                    |            |            |
|----|--------------|-----------------------------------------------------|---------------------|------------|--------------------|------------|------------|
| 44 | but          | sometimes                                           | like Readability is | very       | difficult          | to decide. | (Dina, 6)  |
|    | counter      | force                                               |                     | force      | -cap               |            |            |
| 45 | I'm          | not                                                 | really              | clear      | about Readability. | (Dina, 10) |            |
|    | deny         | force                                               | neg: +cap           |            |                    |            |            |
| 46 | It is        | hard                                                | for me to be        | consistent | you know,          | (Dina, 11) |            |
|    | pronounce    | force                                               |                     | neg: +cap  |                    |            |            |
| 47 | because it's | not                                                 | really              | clear      |                    | (Dina, 12) |            |
|    | deny         |                                                     | force               | neg: +comp |                    |            |            |
| 48 | if           | we have like "this is the criteria of Readability", | that                | easier     | for me             | consistent | (Dina, 13) |
|    | enter        |                                                     | would be            | and be     | about it.          |            |            |
|    | tain         |                                                     | entertain           | +cap       |                    | +cap       |            |

As a result of this lack of clarity, the assessors ended up looking at and marking this feature from somewhat different perspectives. The excerpts below, for example, illustrate how one of the teachers, Dina, used her own discretion in assigning scores to this essay feature.

|    |                                  |                                    |       |                   |            |
|----|----------------------------------|------------------------------------|-------|-------------------|------------|
| 49 | So, I go back to my own criteria | about Cohesion and Readability and | just  | play in that area | (Dina, 36) |
|    | pronounce/inv: +veracity         |                                    | force |                   |            |

|    |         |                              |                                 |                                                  |            |
|----|---------|------------------------------|---------------------------------|--------------------------------------------------|------------|
| 50 | But     | when it comes to readability | I give slightly different score | may be half mark higher and something like that. | (Dina, 37) |
|    | counter |                              | force/ inv: +veracity           |                                                  |            |

The data summarised earlier in Table 6.6 has shown *Vocabulary, Spelling and Punctuation* as the features of students' essays that had one of the highest score variabilities across the six teachers. This confirmed the teachers' perceptions of marking the choices of words as more complicated than marking the other features discussed earlier. Some of the teachers' accounts can be seen in Excerpts 51-54 below:

|    |                                                                              |                                                       |           |
|----|------------------------------------------------------------------------------|-------------------------------------------------------|-----------|
| 51 | OK, the most difficult thing a couple of times, that I feel most of the time | is the vocabulary that the students use in the essay. | (Dina, 1) |
|    | force                                                                        | -cap                                                  | force     |
|    | pronounce                                                                    |                                                       |           |

|    |                           |                            |                                                      |              |
|----|---------------------------|----------------------------|------------------------------------------------------|--------------|
| 52 | Vocabulary, some students | tend to use very big words | without even checking whether it is suitable or not. | (Jelita, 30) |
|    | force                     | entertain                  | force -comp                                          | inv: -cap    |

|    |                         |                            |                                   |                     |
|----|-------------------------|----------------------------|-----------------------------------|---------------------|
| 53 | But some other students | they use simple vocabulary | but their essays are really good. | (Jelita, 31)        |
|    | counter                 | force                      | +comp                             | counter force/+reac |

|    |                 |                              |              |
|----|-----------------|------------------------------|--------------|
| 54 | I find it hard  | in this area for vocabulary. | (Jelita, 32) |
|    | pronounce/ -cap |                              |              |

It was also suggested that *Spelling* and *Punctuation* be separated from *Vocabulary*. As Yusuf stated below, *Vocabulary* should not be put together with *Spelling* and *Punctuation* as one criterion in the rubric. He deemed the former as choice of words.

|    |                                  |                                 |          |             |
|----|----------------------------------|---------------------------------|----------|-------------|
| 55 | Vocabulary, Spelling,Punctuation | I think they could be separated | as well. | (Yusuf, 28) |
|    |                                  | entertain                       | -comp    |             |

|    |           |                                    |                   |                         |
|----|-----------|------------------------------------|-------------------|-------------------------|
| 56 | Probably  | Spelling and Punctuation should be | in the same group | never mind. (Yusuf, 29) |
|    | entertain |                                    | -comp             |                         |

It is interesting to see how Yusuf showed his conceded agreement with the criterion above. Although he perceived *Vocabulary* as a different criterion that should not have been mixed with *Spelling* and *Punctuation*, the last clause of his statement ‘never mind’ concededly agreed with the criteria used. Coming as it did from a teacher employed in the language institution where this study was conducted, this comment functioned to judge the ethical aspect (*propriety* in Appraisal term) of his marking practice in that it showed his solidarity and abidance with the rules stipulated by the institution he worked in. This resonated the overall finding on the ethicality of teachers’ attitudes (*propriety* in Appraisal term) previously summarised (see Section 6.1.2, Table 6.2).

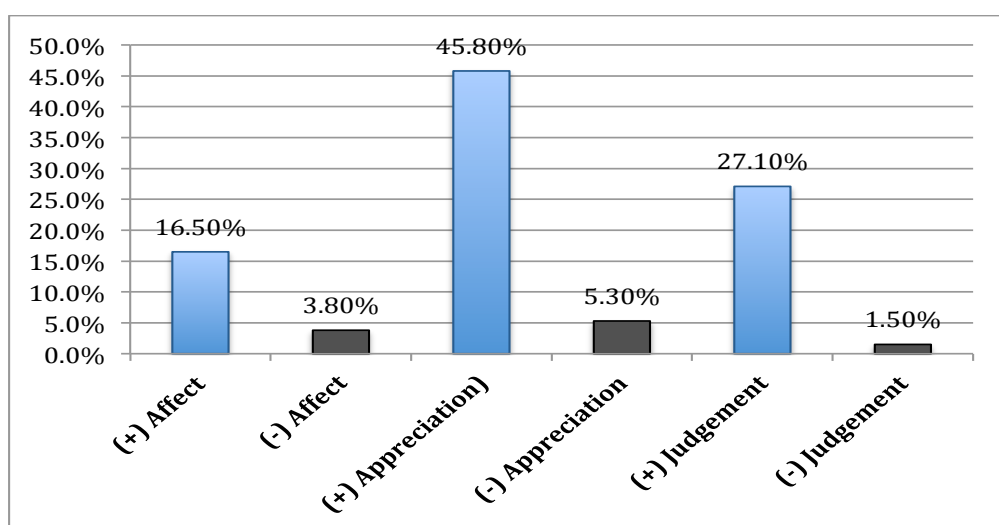
To summarise this section, it can be concluded that the teachers were aware of their practices; that is, they recognised limitations in their marking capabilities, the gap between assessment and teaching practices, and the shortcomings of the existing marking rubric used in terms of choices of criteria and clarity. Overall, teachers’ opinions of their assessment practices were reflected in their actual marking practices. For example, the features of essays teachers found easy to mark, such as *Subject-Verb Agreement*, had a high inter-rater reliability score. On the other hand,

the features they perceived as difficult – *Topic-Supporting Sentences*, *Vocabulary* and *Readability* – had no consensus estimate of inter-rater reliability score, i.e., the discrepancies were more than 1 (one) point higher or lower. This confirms some previous studies such as Hyland (2013) and Matsuda, Saenkum and Accardi (2013). The next section focuses on teachers' opinions and marking practices when using the SFL-informed criteria (post-intervention).

## **6.2 Teachers' perceptions of the SFL-informed rubric**

The first section has described the teachers' perceptions of the existing rubric and identified the sources of inconsistency in their essay marking practices. The findings presented have highlighted the teachers' perceived problems and sources of inconsistency in essay marking. This second section of the chapter answers the last two research questions posed above: RQ3 *How do raters' marking practices shift after using SFL-informed criteria?* and RQ4 *How do raters respond to the SFL-informed criteria?* The purpose of this section is twofold. First, it aims to consider changes in the teachers' perceptions after the professional learning intervention and the adoption of the new rubric, both perceptions of the new rubric as well as of their marking capabilities when using the new rubric. The second purpose is to consider changes in the teachers' marking practices in terms of similarity/dissimilarity of scores assigned among the teachers using the inter-rater reliability tool. Adopting a similar structure to that of the previous section, findings about teachers' perceptions are followed by findings about actual marking practices.

Appraisal analysis reveals patterns of attitudes in the teachers' discourse regarding the new rubric and their marking consistency. The dominant choice made by the teachers in their interviews after the intervention involves positive ATTITUDE resources (Affect, Appreciation and Judgement), indicating that the teachers, in general, perceived the new rubric and their capacities in marking students' essays positively. As seen below (see Figure 6.2), the detailed number of attitudinal items the teachers used when giving opinions about the SFL-informed rubric and their marking practice using the rubric shows that there is a distinctive preference for Appreciation (68) over Judgement (38) and Affect (27) in the six teachers' accounts. This configuration of ATTITUDE resources differs strikingly from the results of the Appraisal analysis of pre-training interviews discussed in the previous section above, in that the number of negative attitudes (Appreciation and Affect) was higher than that of the positive ones. However, this result was similar in terms of the low use of evaluations sourced from feeling or emotion (known as Affect in Appraisal system) in the teachers' responses – a result not surprising given the teachers' professionalism and commitment to their work.



**Figure 6.2 Attitudinal choice: positive and negative perceptions (post-training)**

The high number of positive Appreciation and Affect choices suggests the teachers' overall satisfaction with the rubric used, which in this case had met their expectations. Similarly, the high number of positive Judgement can be seen as the teachers' affirmation of their capabilities in marking using the newly designed rubric. However, a closer examination of the distribution of different kinds of ATTITUDE alongside those of ENGAGEMENT and GRADUATION is warranted to give a more complete picture of the shift in teachers' perceptions.

### **6.2.1 Teachers' opinions of the SFL-informed rubric**

Teachers' perceptions of the new rubric were predominantly positive. These are reflected in the extent to which teachers used Affect and Appreciation when referring to the rubric. As seen in Table 6.8 below, almost 90% of the teachers' talk expressing their opinions of the SFL-informed rubric contained positive attitudes. Of the two Appraisal sub-systems used by teachers to evaluate the rubric, it is apparent that Appreciation was more prevalent than Affect in the distribution of ATTITUDE resources. However, the patterns of teachers' use of these resources to evaluate the new rubric were similar; the proportion of positive comments was higher than the negative ones. As seen below (see Table 6.8), the ratio of positive to negative opinions is 22 to 5 in terms of Affect and 61 to 7 in terms of Appreciation elements. Looking at prevalent use of Appreciation in teachers' talk about the rubric (71.5%), it is evident that teachers' evaluations were 'objective' or considered – that is, they evaluated the rubric as an object (thing) – a tool to improve practice.

**Table 6. 8 Teachers' positive and negative perceptions of the SFL-informed rubric**

| Affect |       |      |      | Appreciation |       |       |       |      |       | Total |
|--------|-------|------|------|--------------|-------|-------|-------|------|-------|-------|
| - sat  | +sat  | -sec | +sec | -reac        | +reac | -comp | +comp | -val | +val  |       |
| 1      | 17    | 4    | 5    | 1            | 2     | 4     | 43    | 2    | 16    | 95    |
| 1.1%   | 17.9% | 4.2% | 5.3% | 1.1%         | 2.1%  | 4.2%  | 45.2% | 2.1% | 16.8% | 100%  |

The prevalence of Appreciation in teachers' talk took the form of direct and explicit evaluations of the SFL-informed rubric, indicating the relative lack of subjectivity in the evaluations made. As we shall see later, the new rubric is appraised as a tool that should be evaluated based on one's perceptions of its 'fit' for purpose e.g., clarity, order, balance, scope and others, rather than based on emotional reactions sourced from personal or emotional feelings e.g., like, love, hate, detest and others. In Excerpt 5, for example, one of the teachers, Ratna, opted to say "The direction *is clearer*" (Appreciation/*composition*) rather than focusing on her personal feeling using an element of Affect/*satisfaction* by saying, "I really *like* the direction in the new rubric." As Martin (1997; 2000) argues, the teachers' preference for Appreciation, as the valuing of something, functions to objectify the evaluation in the discourse. Thus, the teachers' prevalent employment of Appreciation elements in their comments served to make their evaluations become more concrete. From the dialogic perspective, Ratna's use of an unequivocal statement above (*monogloss* in Appraisal term) functioned to show her strong opinion of the clarity of the SFL-informed rubric and did not give any room to argue against it. The intensification used realised by the element of *force* made her positive evaluation of the rubric become even stronger. In this way, the teachers can be seen to be discerning about interventions seeking to improve practice.

As mentioned earlier, teachers' evaluations of the SFL-informed rubric were quite objective in that the evaluations made were mainly aimed at the construct and choice of criteria (*composition* in Appraisal term) of the new rubric. In fact, almost half (49.5%) of the teachers' opinions was about the compositional aspects of the new rubric, and these opinions were predominantly positive. As summarised in Table 6.8 above, with a figure of 45.2%, the proportion of positive *composition* identified in the teachers' accounts was more than ten times higher than that of the negative ones (4.2%). As *composition* is concerned with views of order, balance or complexity (Martin and White 2005, White 2002), its prevalence can be seen as representation of teachers' satisfaction with the design and structure of the SFL-informed rubric used. The affirmation of this interpretation can be seen in excerpts below. In Excerpts 1-5, for example, teachers positively evaluated the SFL-informed rubric as being *clear* or *easy to understand*.

|   |                  |                     |                                                          |                                                     |
|---|------------------|---------------------|----------------------------------------------------------|-----------------------------------------------------|
| 1 | Because it is    | clear               | over there what you have to assess and what to look for. | (Dina, 19)                                          |
|   | mono             | +comp               |                                                          |                                                     |
| 2 | It is            | clear.              |                                                          | (Dina, 36)                                          |
|   | mono             | +comp               |                                                          |                                                     |
| 3 | The direction is | clearer.            |                                                          | (Ratna, 5)                                          |
|   | mono             | force/+comp         |                                                          |                                                     |
| 4 | The features are | easy to understand. |                                                          | (Mirzan, 14)                                        |
|   | mono             | +comp               |                                                          |                                                     |
| 5 | I think          | they are            | quite clear                                              | for me to understand the criteria here. (Yusuf, 10) |
|   | entertain        |                     | force +comp                                              |                                                     |

Similarly, as shown in Excerpts 6-9, the rubric was perceived as more comprehensive, more detailed or more complete, indicating that the teachers do not shy away from complicity in the professional tools they employ:

|   |                                                                                    |  |       |  |       |  |
|---|------------------------------------------------------------------------------------|--|-------|--|-------|--|
| 6 | Because the scheme and criteria <b>are</b> <b>more</b> <b>detailed</b> (Jelita, 2) |  |       |  |       |  |
|   | mono                                                                               |  | force |  | +comp |  |

|   |                                                                                                                                   |  |       |  |       |       |
|---|-----------------------------------------------------------------------------------------------------------------------------------|--|-------|--|-------|-------|
| 7 | The marking rubric <b>much more</b> <b>useful</b> because <b>more</b> <b>complete</b> . (Mirzan, 2)<br>given to me <b>is</b> it's |  |       |  |       |       |
|   | mono                                                                                                                              |  | force |  | +val  |       |
|   |                                                                                                                                   |  |       |  | force | +comp |

|   |                                                                                          |  |       |  |       |  |
|---|------------------------------------------------------------------------------------------|--|-------|--|-------|--|
| 8 | So, it <b>is</b> <b>more</b> <b>complete</b> , <b>more</b> <b>detailed</b> . (Mirzan, 3) |  |       |  |       |  |
|   | mono                                                                                     |  | force |  | +comp |  |

|   |                                                               |  |       |  |       |  |
|---|---------------------------------------------------------------|--|-------|--|-------|--|
| 9 | This <b>is</b> <b>more</b> <b>comprehensive</b> . (Yusuf, 19) |  |       |  |       |  |
|   | mono                                                          |  | force |  | +comp |  |

Excerpts 10-13 demonstrate that teachers evaluated the construct and choices of criteria in the rubric positively.

|    |                                                                                                              |  |      |  |       |  |
|----|--------------------------------------------------------------------------------------------------------------|--|------|--|-------|--|
| 10 | The <b>breakdown of the marking criteria</b> <b>allowed</b> me to add <b>more substantively</b> (Whitney, 2) |  |      |  |       |  |
|    | t: +comp                                                                                                     |  | mono |  | force |  |

|    |                                                                                                                                                                    |  |          |  |  |  |
|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|----------|--|--|--|
| 11 | <b>But</b> these marking criteria <b>definitely give me space to assess students' essay content not just grammatical inaccuracy and things like that.</b> (Dina,8) |  |          |  |  |  |
|    | counter                                                                                                                                                            |  | t: +comp |  |  |  |

|    |                                                                                             |  |          |  |  |  |
|----|---------------------------------------------------------------------------------------------|--|----------|--|--|--|
| 12 | <b>Yeah,</b> in the new system (rubric) <b>the content is also evaluated</b> . (Jelita, 14) |  |          |  |  |  |
|    | concur                                                                                      |  | t: +comp |  |  |  |

|    |                                                                                      |  |          |  |  |  |
|----|--------------------------------------------------------------------------------------|--|----------|--|--|--|
| 13 | <b>OK, like I said, it DID</b> <b>make it easier to be consistent</b> . (Whitney, 3) |  |          |  |  |  |
|    | pronounce/ force                                                                     |  | t: +comp |  |  |  |

The teachers' positive opinions about the construct and choice of criteria in the new rubric were particularly strong. As seen in Excerpts 1-13 above, the majority of the propositions were uttered in a form of bare assertions (*monogloss* in Appraisal terms) and were formulated in such a way that did not allow or provide any space for other alternate opinions. Taking Martin and White's (2005) perspective, the teachers' bare assertions shows their unequivocal opinions of the rubric. Thus, when

the teachers, such as Dina, Ratna and Mirzan said, “It **is** clear”, “The direction **is** clearer” and “The features **are** easy to understand” and as in Excerpts 2, 3 and 4, they considered such statements as a “truth” or “fact” about the new rubric. The use of GRADUATION resource of *force/intensification* such as *more detailed*, *more complete*, *clearer*, and *more comprehensive* as seen in the excerpts above, functioned to strengthen the teachers’ opinions about the compositional aspects of the rubric in the propositions made. It might be argued that the prevalence of *force/intensification* may have to do with how the questions were asked. In other words, the interview questions might have prompted the teachers to give responses in certain ways. However, this was not the case. In the interview question no. 2 “Now, I want to ask your opinions about the new marking rubric; how do you find the new marking criteria?” the question was open-ended and did not prompt the respondents to use the evaluative vocabulary quoted above (see Appendix C for the detailed interview guides of the study). As seen in the excerpts above, isolation realisations such as *clearer*, *more comprehensive*, *more complete*, *more detailed* functioned to intensify the degree of clarity of the new rubric.

Another noteworthy finding showing teachers’ positive opinions of the new rubric was the dominance of positive *valuation* in the teachers’ responses. Of the eighteen *valuation* items identified, only two negative values, such as *most difficult* and *a bit abstract*, were found in the teachers’ interviews, while the other sixteen items all contained positive values such as *helpful*, *help*, *useful* and *important*. As a system of meaning focusing on considered opinion or the estimation of the worth of something (Martin and White 2005; 2007), which in this case is the SFL-informed rubric, the prevalence of positive *valuation*, suggesting the teachers’ perceptions of the overall

worthwhileness of the new rubric used. As seen in Excerpts 14-19 below, the new rubric was regarded as something helpful and useful in the teachers' marking practice. When we see them from the degree of evaluative words used (GRADUATION in Appraisal term), the worthwhileness of the new rubric was represented by predominant use of the evaluative words using the elements of *force/intensification*. These positive evaluations become even stronger when the use of ENGAGEMENT resources is considered, as the majority of the statements were made as *pronouncements*. *Pronounce* functions to show authorial emphases in the statements made by speakers or writers (White 2001; 2012). Likewise, the prevalent instances of *pronouncement* formulations in the teachers' accounts functioned to show their investment in the statements made.

|    |                                                                                                             |       |       |       |       |       |
|----|-------------------------------------------------------------------------------------------------------------|-------|-------|-------|-------|-------|
| 14 | It <b>certainly</b> <b>helps</b> me to be <b>more</b> <b>consistent</b> . (Dina, 1)                         |       |       |       |       |       |
|    | pronounce                                                                                                   | +val  |       | force | +cap  |       |
| 15 | The marking rubric <b>much more</b> <b>useful</b> because <b>more</b> <b>complete</b> . (Mirzan, 2)         |       |       |       |       |       |
|    | mono                                                                                                        | force | +val  |       | force | +comp |
| 16 | and it is <b>very</b> <b>important</b> . (Jelita, 15)                                                       |       |       |       |       |       |
|    | mono                                                                                                        | force | +val  |       |       |       |
| 17 | and it <b>helps me also</b> to look at the aspects of the essay <b>more</b> <b>consistently</b> (Ratna, 27) |       |       |       |       |       |
|    | +val/pronounce                                                                                              |       |       | force | +cap  |       |
| 18 | <b>Well so far, I find it</b> <b>quite</b> <b>helpful</b> . (Yusuf, 9)                                      |       |       |       |       |       |
|    | pronounce                                                                                                   | force | +val  |       |       |       |
| 19 | <b>Appraisal yeah</b> that was <b>very</b> <b>helpful</b> . (Whitney, 8)                                    |       |       |       |       |       |
|    | pronounce                                                                                                   |       | force | +val  |       |       |

The teachers' endorsement of the rubric was strong. Based on the close analysis of Affect, as summarised earlier in Table 6.8 above, it became evident that the majority of teachers' talk reported their positive assessment in the new rubric. Of the 18 elements of Affect identified, almost all of them (17 items) reflected the teachers'

*satisfied* feelings towards the rubric. This finding suggests the teachers preferred to project their perceptions of the SFL-informed rubric through feelings of contentment rather than feelings of confidence. As a sub-system of Affect concerned with feelings of achievement and frustration (Martin and White 2005, White 2001), the overtly high number of instances of positive *satisfaction* in teachers' statements suggested the absolute feelings of capability among teachers in using the new rubric. When seen from a dialogic perspective, the teachers' statements were predominantly made as *pronouncements*. This might be because the teachers wanted to add emphasis to their authorial, or the 'I', voice, which realised their value position but at the same time, invited other voices. In other words, although the teachers' convictions about the rubric were strong, they were aware that the opinions could differ. That is why the statements made, such as "It's definitely I could say my favourite ...", "I really like the breakdown and "I especially also like ..." (as in Excerpts 20-22), were focused on the speakers' personal notes.

|    |                                                           |      |  |  |
|----|-----------------------------------------------------------|------|--|--|
| 20 | It's definitely I could say my favourite part. (Dina, 18) |      |  |  |
|    | pronounce                                                 | +sat |  |  |

|    |                                                                  |      |  |  |
|----|------------------------------------------------------------------|------|--|--|
| 21 | I really like the breakdown, especially with genre. (Whitney, 5) |      |  |  |
|    | force                                                            | +sat |  |  |
|    | pronounce                                                        |      |  |  |

|    |                                                                        |      |  |  |
|----|------------------------------------------------------------------------|------|--|--|
| 22 | I especially also like the communication roles, the tenor (Whitney, 7) |      |  |  |
|    |                                                                        | +sat |  |  |
|    | pronounce                                                              |      |  |  |

Note that the word *favourite* and *like* in Excerpts 23-24 were treated as *satisfaction* as its meaning was more about the speaker's contentment or feeling of achievement with the new rubric than her feeling of happiness.

The teachers did, however, express some anxiety or insecurity about specific criteria in the new rubric. The excerpts below (see Excerpts 23 and 24) were Dina's response to questions about the most difficult and the easiest features of student essays to mark when using criteria in the new assessment rubric.

|    |     |                 |          |                        |            |
|----|-----|-----------------|----------|------------------------|------------|
| 23 | I'm | slightly, a bit | confused | with Textual Cohesion. | (Dina, 12) |
|    |     | force           | -sec     |                        |            |
|    |     | pronounce       |          |                        |            |

|    |            |                                                              |            |
|----|------------|--------------------------------------------------------------|------------|
| 24 | Definitely | the Genre, the social purpose and stages developed in essays | (Dina, 17) |
|    | +sec/      | pronounce                                                    |            |

In terms of polarity of *in/security* elements in the teachers' accounts, as shown earlier in Table 6.8, the positive and negative *in/security* features employed were almost similar in number. Of all the nine items of *in/security* identified in teachers' texts, five elements were positive or *security* while the remaining four were negative or *insecurity*. Despite the slightly higher number of positive perceptions (*security*), which may be seen as an indication of the teachers' confident feelings when using the new marking rubric, the similar number of negative perceptions suggested the opposite effects to the teachers' confidence in using the SFL-informed rubric used. However, when seen from the evaluative words used, teachers' positive or secure feelings when using the SFL-informed rubric were stronger than the negative ones.

In terms of intensity, the teachers' negative or insecure feelings were weaker than the positive ones. In Excerpt 23, Dina, one of only two teachers whose comments consisted of *in/security* elements, expressed her insecure feeling through her use of the word *confused*, but at the same time down-toned it by using two elements of *force* (*slightly* and *a bit*). This created a less intense negative feeling. In contrast, her use of *definitely* in Excerpt 24 created a higher intensity of positive or secure feeling

at marking the genre stages of student essays. Note that although *definitely* in Excerpt 26 can be seen as infused GRADUATION (Martin and White 2005) i.e., it belongs to the element of *force*, it was regarded as an element of ATTITUDE (Affect) when related to the interview question: *All right. You've told me that the SFL-informed rubric makes it easier for you to mark student essays. Now, can you tell me the features of the essays you find easiest to mark then?* The interview question makes it clearer for us to see that *definitely* here was more related to the teacher-confident or secure feeling to mark some features of the essays using the new rubric, which was intensified by a *force* element. This positive or secure feeling was intensified further by her use of *favourite (satisfaction)* and *clear (composition and invoked satisfaction)*. This suggests that the teacher was, overall, satisfied with the new rubric and her capacity to use it.

In summary, teachers drew on the resources of Appreciation to describe the SFL-informed rubric, identifying features such as *Genre*, *Content* and *Communication roles (tenor)* as important features of the rubric. They pointed out that the SFL-informed rubric is more detailed and systematic in the choice of criteria, and thus is perceived as 'helpful' and 'useful' in promoting the reliability of their marking practices. The analysis also revealed something of the teachers' anxiety about their practices and use of the rubric. The next section focuses on analysing teachers' judgements of their consistencies after the intervention.

### 6.2.2 Teachers' opinions about their consistencies and practices when using the SFL-informed rubric

Teachers' opinions of their consistencies in marking after the intervention were predominantly positive. This can be seen by the pervasive employment of positive Judgement of their own *capacity* in the interview data. As seen in Table 6.9, positive *capacity* constituted almost 74% of all the Judgement elements found in teachers' accounts. Of all the 38-Judgement items found, 30 items (79%) belonged to *capacity* while the rest were shared among *propriety* (13.2%), *normality* (5.3%) and *veracity* (2.6%). The prevalence of *capacity* indicated the teachers' perceptions of their reliabilities and marking practices, and were mainly centred on capabilities, i.e., their professional skills and knowledge. This is in contrast to the result of the pre-training analysis, in which *capacity* constituted only 34% of all the sub-system of Judgement.

**Table 6. 9 Teachers' use of Judgement in terms of capabilities and practices when using the SFL-informed rubric**

| Judgement |       |      |       |      |      |      |      |       |       | Total |
|-----------|-------|------|-------|------|------|------|------|-------|-------|-------|
| -norm     | +norm | -cap | +cap  | -ten | +ten | -ver | +ver | -prop | +prop |       |
| -         | 2     | 2    | 28    | -    | -    | -    | 1    | -     | 5     | 38    |
| -         | 5.3%  | 5.3% | 73.6% | -    | -    | -    | 2.6% | -     | 13.2% | 100%  |

The pervasive employment of capacity, such as in Excerpts 25-30, can be seen as strong evidence of the teachers' affirmation of heightened capability to mark student essays using the new rubric. This affirmation was clearly expressed by their use of choices of inscribed Judgements such as *confident* and *consistent*, intensified by elements of *force* (e.g., *more*). This improved capability was also affirmed by the use of negations of negative *capacity*, such as *cannot be subjective*.

|    |                                                                                                    |       |           |       |      |  |
|----|----------------------------------------------------------------------------------------------------|-------|-----------|-------|------|--|
| 25 | It <b>certainly</b> <b>helps</b> me to be <b>more</b> <b>consistent</b> . (Dina, 1)                |       |           |       |      |  |
|    | pronounce                                                                                          | +val  |           | force | +cap |  |
| 26 | I <b>find myself</b> <b>more</b> <b>consistent</b> after the training. (Jelita, 1)                 |       |           |       |      |  |
|    | pronounce                                                                                          | force | +cap      |       |      |  |
| 27 | So, I <b>cannot</b> be <b>subjective</b> anymore. (Mirzan, 4)                                      |       |           |       |      |  |
|    |                                                                                                    | deny  | neg: +cap |       |      |  |
| 28 | I <b>think</b> it's <b>more</b> <b>consistent</b> . (Ratna, 1)                                     |       |           |       |      |  |
|    | entertain                                                                                          |       | force     | +cap  |      |  |
| 29 | The way I assess the students' work is going <b>more</b> <b>consistent</b> than before. (Yusuf, 2) |       |           |       |      |  |
|    | <b>to be</b>                                                                                       |       |           |       |      |  |
|    | pronounce                                                                                          |       |           | force | +cap |  |
| 30 | I <b>find it</b> <b>more</b> <b>consistent</b> . (Whitney, 1)                                      |       |           |       |      |  |
|    | pronounce                                                                                          | force | +cap      |       |      |  |

When assessing the ethical dimensions of their practices, the teachers also perceived their actions positively. This can be seen by the predominance of positive *propriety* in the teachers' talks (see Table 6.9 above). Of all the five *propriety* items identified, none was found to carry negative value. Due to the fact that the teachers had used the SFL-informed rubric for a relatively short time (approximately two weeks), the small number of evaluations which focused on the appropriateness of conducting the assessment (*propriety* in Appraisal terms) was not unexpected.

An interesting finding here was the fact that the *propriety* elements used were all in the forms of invocation, that is, made without using any attitudinal words (Martin and White 2005). In the two excerpts below, the selections of ideational meaning (expressing ideas or experiences about the world through clause constructions) "while before I didn't care" and "but also assess the content" were enough to imply ethical positions of the teachers' marking practices. In other words, the teachers were making judgements reflecting on their earlier marking practices using the

existing, non-SFL rubric. As seen in Excerpts 31 and 32, Ratna and Dina reflect on their past marking practices to show contrast to what they did after the training. The use of an adverb (*before*) preceded by conjunctions such as *while* and *but* functioned to show contrast, and at the same time, invoked judgement of their ethicality when marking using the new rubric.

|    |         |          |        |            |                                |             |
|----|---------|----------|--------|------------|--------------------------------|-------------|
| 31 | while   | before I | didn't | care       | as long as the word was right, | (Ratna, 15) |
|    | counter |          | deny   | neg: +prop |                                |             |

|    |     |             |                                      |  |  |           |
|----|-----|-------------|--------------------------------------|--|--|-----------|
| 32 | but | also assess | the content of the students' essays. |  |  | (Dina, 4) |
|    |     | t: +prop    |                                      |  |  |           |

Teachers used Judgements of normality and truthfulness less prevalently. The infrequent employment of these two Judgement sub-systems was quite normal remembering the fact that the teachers' had used the SFL-informed rubric on only one occasion for the duration of approximately one week. Thus, the sparing employment of *veracity* and *normality* in their statements was not unforeseen. Similarly, the prevalence of *normality* in the teachers' accounts when evaluating their practices using the existing marking rubric, as previously discussed in Section 6.1.2, was not unexpected.

## 6.2.3 Teachers' opinions of the intervention

Table 6.10 Teachers' positive and negative perceptions of the intervention

| Affect |       |      |      | Appreciation |       |       |       |      |       | Total |
|--------|-------|------|------|--------------|-------|-------|-------|------|-------|-------|
| - sat  | +sat  | -sec | +sec | -reac        | +reac | -comp | +comp | -val | +val  |       |
| 1      | 4     | 2    | -    | -            | 2     | 3     | 6     | 1    | 10    | 29    |
| 3.4%   | 13.8% | 6.8% | -    | -            | 6.8%  | 10.3% | 20.7% | 3.4% | 34.5% | 100%  |

The teachers offered useful evaluations of the intervention. Across the statements about the intervention, teachers' opinions of the program were predominantly

positive. Of the 29 evaluative words identified, 22 items (76%) were coded as positive evaluations such as “The training itself is **really good**”, “It **helps** me very much”, “Personally, I think the training is good enough” or “The seminar on identifying the choice of words was **very helpful**”. The remaining seven items, which constituted 24% of teachers’ comments about the intervention, had negative connotations such as “well, maybe next time, perhaps more extensive”, “Well, bring participants more samples related to the discussions, to the topics would be a better idea” or “Yeah more exercises probably, because it’s tricky sometimes”.

Most of their quotes are identified as explicit (inscribed in Appraisal terms) Appreciation, which means the teachers targeted their talk to the intervention rather than to the presenter. This shows the directness of the teachers in giving their opinions about the intervention, which they referred to as *good* and *helpful*. These opinions were made stronger by the prevalent employment of *force* elements such as *a lot*, *really*, *very* and *very much* among others, which functioned to strengthen the quality of their evaluative words. As far as the positioning of the teachers’ directness is concerned (ENGAGEMENT in Appraisal terms), some of the quotes were marked with more *monoglossic* statements (see Excerpts 33-36), indicating the teachers’ unequivocal opinions about the worthwhileness of the training program.

|    |                                                    |       |       |              |
|----|----------------------------------------------------|-------|-------|--------------|
| 33 | The training itself is <b>really</b> <b>good</b> , |       |       | (Jelita, 17) |
|    | mono                                               | force | +reac |              |

|    |                                 |  |       |              |
|----|---------------------------------|--|-------|--------------|
| 34 | It is <b>really</b> <b>good</b> |  |       | (Jelita, 18) |
|    | mono                            |  | +reac |              |

|    |                                |            |  |  |
|----|--------------------------------|------------|--|--|
| 35 | It <b>helps me very much</b> . | (Ratna, 3) |  |  |
|    | +val                           | force      |  |  |
|    | mono                           |            |  |  |

|    |                                                                                        |               |      |  |
|----|----------------------------------------------------------------------------------------|---------------|------|--|
| 36 | The seminar on identifying the choice of words <b>was</b> <b>very</b> <b>helpful</b> . | (Whitney, 10) |      |  |
|    | mono                                                                                   | force         | +val |  |

Other responses (see Excerpts 37-39) were less direct, selecting *heteroglossic/pronounce*, which, in the Appraisal framework, indicates explicit authorial emphases of the assertions made (White 2001; 2012).

|    |                                                                |           |  |  |
|----|----------------------------------------------------------------|-----------|--|--|
| 37 | <b>So it certainly</b> <b>answers my question about that</b> . | (Dina, 5) |  |  |
|    | pronounce                                                      | t: +val   |  |  |

|    |                                                                         |           |       |  |
|----|-------------------------------------------------------------------------|-----------|-------|--|
| 38 | <b>And about consistency, definitely</b> <b>it helps</b> <b>a lot</b> . | (Dina, 6) |       |  |
|    | pronounce                                                               | +val      | force |  |

|    |                                                                                                                                                                       |            |        |  |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|--------|--|
| 39 | <b>I learned from</b> marking students' <b>simply</b> <b>grammar, spelling, punctuation,</b><br><b>the training that</b> work is not <b>topic sentence and so on.</b> | (Yusuf, 1) |        |  |
|    | pronounce/+cap                                                                                                                                                        | force      | t:-val |  |
|    |                                                                                                                                                                       | t: +val    |        |  |

Although the teachers found the intervention helpful, they realised that more time should have been allocated to the training. This can be seen from a few negative comments about the structure of the intervention (*composition* in Appraisal terms) (see Extracts 40, 41 and 42 below). The three negative opinions identified in teachers' comments about the structure of the training program suggest that the teachers expected the program to be made more extensive, with more practice on criteria-identification activities.

|    |                                       |            |  |  |
|----|---------------------------------------|------------|--|--|
| 40 | <b>Definitely</b> <b>the timing</b> . | (Dina, 39) |  |  |
|    | pronounce                             | -comp      |  |  |

|    |                                                                                                       |            |  |  |
|----|-------------------------------------------------------------------------------------------------------|------------|--|--|
| 41 | I mean <b>if we can</b> add more, I mean <b>if the session</b> like <b>is not just only two hours</b> | (Dina, 40) |  |  |
|    | entertain                                                                                             | t: -comp   |  |  |

|    |             |          |           |       |            |              |
|----|-------------|----------|-----------|-------|------------|--------------|
| 42 | Well, maybe | the time | perhaps   | more  | extensive. | (Mirzan, 17) |
|    | entertain   | -comp    | entertain | force | -comp      |              |

Such comments tended to be implicit, using the *heteroglossic* choice of *expand/entertain* such as *maybe*, *perhaps* and conditional sentences such as *if we can* and *if the session* to talk about the intervention. *Entertain* in the system of ENGAGEMENT is related to wordings by which the position of the speaker/writer is but one of a number of possible positions, and thereby makes dialogic space for those possibilities (Martin and White 2005). The predominance of this element in teachers' accounts illustrates how the teachers use language to leave the statements open to other opinions, which indicates uncertainty to some extent or their unwillingness to criticise the program.

In summary, it can be concluded that teachers were, on the whole, satisfied with the intervention. This is shown by the predominance of positive evaluations such as *really good*, *helpful* and *helps me very much* identified in the analysis of teachers' comments about the intervention. Teachers' use of *monoglossic* assertions suggests their unequivocal opinions of the usefulness or helpfulness of the training program. There were some responses indicating shortcomings in the training program in terms of the timing (see Extracts 40-42), indicating future considerations for such a program. However, the number was quite insignificant. This suggests that, on the whole, the intervention was conducted successfully (see Excerpts 43 and 44).

|    |           |                       |     |                 |               |
|----|-----------|-----------------------|-----|-----------------|---------------|
| 43 | It seems  | in general everything | was | well presented. | (Whitney, 20) |
|    | entertain | force                 |     | +comp           |               |

|    |           |                                     |               |
|----|-----------|-------------------------------------|---------------|
| 44 | I think   | you did a good job presenting them. | (Whitney, 22) |
|    | entertain | +cap                                |               |

The following section examines whether the teachers' opinions and attitudes are reflected in their real marking capabilities, indicated by their level of consensus estimate of inter-rater reliability (Stemler 2004).

#### 6.2.4 Post-training inter-rater reliability

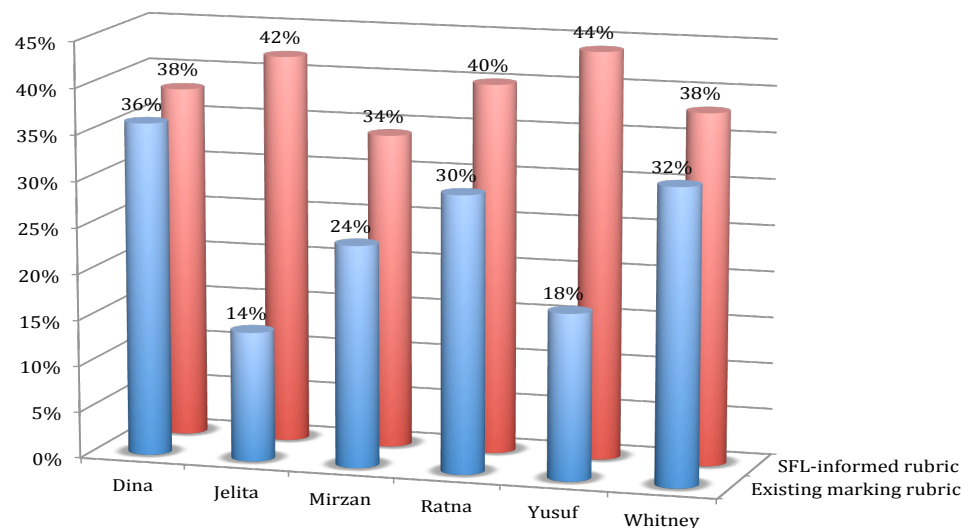
Analysis of marking consistencies shows quite a strong improvement in the teachers' inter-rater reliability rate. As seen in Table 6.11 below, the average consistency rate among the six assessors increased from 25.6% when marking using the existing marking rubric (before the intervention) to 39.3% when marking using the SFL-informed rubric (after the intervention). The assessors with the highest consistency after the training were Yusuf (44%), Jelita (42%) and Ratna (40%). This means that after the training, they were able to reach higher agreement, that is, four out of ten essays they marked had scores similar to those assigned by other teachers. Dina and Whitney had similar consistency rate (with 38% each) while Mirzan, with 34% agreed scores, had the lowest consistency rate among them. In other words, on average, Mirzan had three and a half essays that received scores similar to those assigned by his colleagues. On the whole, the teachers had 20 essays with similar scores out of a maximum of 50 essays, with 40% mean consensus estimate of inter-rater reliability.

**Table 6.11 Post-training consensus estimate of inter-rater reliability**

| <b>Raters</b> | <b>Similar Score Essay</b> | <b>Percentage</b> |
|---------------|----------------------------|-------------------|
| Dina          | 19/50                      | 38%               |
| Jelita        | 21/50                      | 42%               |
| Mirzan        | 17/50                      | 34%               |
| Ratna         | 20/50                      | 40%               |
| Yusuf         | 22/50                      | 44%               |
| Whitney       | 19/50                      | 38%               |
| <b>Mean</b>   | <b>118/300</b>             | <b>39.3%</b>      |

Strong evidence of improvement was found in teachers' marking of students' essays using the new rubric, both collectively and individually. Collectively, there was approximately a 14% increase in the mean consensus estimate of inter-rater reliability when using the SFL-informed rubric. This increase was more apparent when seen on an individual basis. On an individual basis, Jelita and Yusuf made the most significant improvement in inter-rater reliability scores, with 28% and 26% improvement rates respectively. Before the training, Jelita and Yusuf were the least consistent assessors, with average inter-rater reliability scores of 14% and 18% respectively (see Figure 6.3 below). That is, merely less than two essays received scores similar to those assigned by other teachers. However, after the intervention, they both had the highest rates with 42% and 44% respectively. Ratna demonstrates a 10% improvement i.e., from 30% inter-rater consistency rate using the existing marking rubric to 40 % when using the SFL-informed rubric. Whitney and Dina had similar inter-rater reliability rates but varied in the improvement made. Dina, who had the highest inter-rater reliability score when using the existing marking rubric, made only 2% improvement when using the SFL-informed rubric (36% to 38%) while Whitney made a slightly better improvement in her marking consistency (from 32% to 38% after the intervention). Although Mirzan had the lowest consistency

rate among the teachers, with 6%, he made slightly higher improvement than Dina (24% to 34%).



**Figure 6.3 Teachers' consistency before and after the training**

Of the ten essays marked using the SFL-informed rubric, Essays 8a and 10a were found to have the highest frequency (see Table 6.12). They respectively received 22 and 16 similar scores out of a maximum of 30 recurrences. Features with the most and least inter-assessor agreement in the two essays were then analysed so as to give a clear picture of shifts in the teachers' marking practice and to identify which still caused problems with consistency (see RQ3 above). In order to set the features, it was necessary to classify the genre of this essay so that the scores given and the conformity of the essay to stages and phases of a particular genre could be identified.

Table 6.12 Classification of student essays: genre, mean score and frequency

| Topic: Globalisation                                                        |                       |            |           |
|-----------------------------------------------------------------------------|-----------------------|------------|-----------|
| Task Prompt: Globalisation can have negative impact on developing countries |                       |            |           |
| Essay                                                                       | Genre                 | Mean score | Frequency |
| 1a                                                                          | Analytical exposition | 13.1       | 8         |
| 2a                                                                          | Analytical exposition | 14.3       | 12        |
| 3a                                                                          | Analytical exposition | 14.8       | 10        |
| 4a                                                                          | Analytical exposition | 13.8       | 8         |
| 5a                                                                          | Analytical exposition | 13.3       | 12        |
| 6a                                                                          | Analytical exposition | 10.7       | 14        |
| 7a                                                                          | Analytical exposition | 13.8       | 10        |
| 8a                                                                          | Analytical discussion | 16.7       | 22        |
| 9a                                                                          | Analytical discussion | 9.1        | 6         |
| 10a                                                                         | Analytical exposition | 9.6        | 16        |

Based on the classification of the ten essays in terms of their genre (see Table 6.12 above) eight essays were identified as analytical exposition and two essays (Essays 8a and 9a) as analytical discussion. As explicated previously in Chapter Four, the essay with highest frequency of consistency was analysed to find out the correlation between the scores the teachers assigned to the stages and/or phases identified and its conformity with the stages of an argument essay. In addition to genre stages, the marks assigned to the highest-scored essay were analysed in terms of register variables such as topic research and content (*field*), communication roles (*tenor*) and written features (*mode*). The last part of analysis of essay scores concerned features which showed linguistic accuracy such as *grammar*, *spelling*, *punctuation* and *word limit* of the essay (see Appendix E for a sample of the SFL-informed rubric used).

The data above (see Table 6.12) has shown the compatibility of Essay 8a and Essay 10a among the six teachers in terms of frequency. This means that these essays had the highest agreement among the six assessors. The two essays, however, differed significantly in terms of the assigned scores. Essay 8a received the highest mean score of 16.7 out of a maximum of 20 while Essay 10a had the second-lowest score with an average of 9.6 out of a maximum of 20, or 4.8 on a ten-point scale. Similar

to the analysis of pre-training essays, two essays with the highest agreement among teachers were analysed in this part of the study.

Analysis of genre stages (see Appendix F) revealed Essay 8a as a typical instance of Analytical Discussion. It had the distinctive quality of a Discussion in that it conformed and showed control over the generic structure of the essay, i.e., **Issue – Sides – Resolution** as seen in Coffin (2004) and Rose and Martin (2012), as well as the phases previously discussed in the previous chapter (see Chapter Three, Figure 3.4 for stages and phases of the Discussion used).

Despite its conformity and control of the generic structure, this essay was slightly lacking clarity in the **preview** phase when directing its reader(s) on how the essay would develop. In this phase the writer's statement of: *This argument essay will explain more about the impacts of globalisation on developing countries* was quite acceptable but it did not give a clear indication of the kinds of impact it would explain in the body paragraph. As the **preview** functions to predict what each subsequent argument or side will be about (Rose and Martin 2012), the writer should have stated it more clearly to signal his approach to the **Issue**. Taking the size of the essay with a maximum of 300 words into consideration, perhaps adding quality words or adjectives such as *positive* and *negative* to appraise the effects of globalisation as in: *This argument essay will explain more about the positive and negative impacts of globalisation on developing countries*, would have been sufficient to preview the **Sides**. In the new rubric, this was included in the second section of the criterion of Genre: social purpose and stages developed in essays: *The*

*first paragraph provides preview of position and main themes through which this position/argument will develop (see Appendix E).*

In the body paragraph, the *hyperThemes* or **topics** were quite simple but well formulated in that it stated the main ideas of the paragraphs straightforwardly. The statement of *Globalisation has given several advantages to developing countries* in the first body paragraph (**Side 1**), for example, was sufficient to highlight the whole content of the paragraph, as to how the *hyperTheme* or **topic** would be developed. Similarly, the writer's statement of: *Despite all the positive impacts mentioned above, globalisation has affected developing countries in negative ways*, as seen in the second body paragraph (**Side 2**), was considered effective. It signalled the readers a transitional position with a conjunction 'despite' to signpost the opposite view, and it used an adjective or Appreciation (e.g., *negative ways*) to effectively support the preceded conjunction to highlight what the paragraph would be about. In the new rubric, this was addressed in the third section of Genre: *Develops main themes and their relationships in a series of well organised paragraphs which follow the order given in the first paragraph are logically and clearly organised* (see Appendix E).

The writer's statements in the **Resolution** stage were well communicated. The sentence *In conclusion, globalisation can cause positive and negative impacts on developing countries, but the negative impacts are outweighing the positive impacts* in the **review** phase, distilled the points that the writer had made in the body paragraphs quite succinctly. The sentence: *It causes the exploitation of natural resources and labour work* in the **resolve** phase reiterated the negative effects of

globalisation as had been argued in the **Side 2** stage. However, the writer's use of a linker in addition, as in *In addition, the benefit of globalisation is not universal. The rich are getting richer and the poor can get a better living but they are not going anywhere*, was seen as counter-effective because it signalled the writer's introduction of a new idea. In the redesigned rubric (see Appendix E), this criterion was included in the last section of Genre: *Concludes with overall interpretation, summarising the themes of the topic and highlighting their significance in light of the approach taken in the subject.*

It can be seen from the data in Table 6.13 that the score gap for each stage and phase of Essay 8a was relatively small, and according to Stemler's (2004) formula, was compatible in terms of consensus estimate of inter-rater reliability counts. In the **issue** phase, for example, the scores the six teachers assigned were exactly the same i.e., on a scale of 0-3 (zero to three), all six teachers assigned the highest possible score, 3 (three) to this phase of Essay 8a. This, according to Rose's (2005) score traits, means that the teachers agreed that the student's writing meets the criterion at the highest level expected as explicated in the rubric (see Chapter Five, Table 5.6 for the descriptive information of each score trait). In the **preview**, the scores the teachers gave to this phase of Essay 8a were similar in that the gap between the highest and lowest score was relatively small, i.e., the highest score was 3 (three) and the lowest was 2 (two).

Table 6.13 Score gap of Essay 8a and Essay 10a based on genre stages

| Essay 8a      |       |       |            | Essay 10a |          |             |
|---------------|-------|-------|------------|-----------|----------|-------------|
| Stages        | Issue | Sides | Resolution | Thesis    | Argument | Restatement |
| Lowest score  | 2.5   | 2     | 3          | 1         | 1        | 0           |
| Highest score | 3     | 3     | 4          | 2         | 2        | 2           |
| Score gap     | 0.5   | 1     | 1          | 1         | 1        | 2           |

Essay 10a (see Appendix F), another text that markers agreed upon, is a typical instance of Analytical Exposition but demonstrates a lack of control of the generic structure of the essay, i.e., **Thesis – Argument – Restatement** as seen in Martin (1985), Rose (2005) and Rose and Martin (2012). In the **Thesis** stage, for example, the **position** of the writer – *Today developing countries can be impacted by globalisation, and the impact can be negative* – was not clear and gave the impression of a somewhat neutral position, as if the writer would like to weigh both positive and negative effects of globalisation on developing countries. In the **preview** phase, however, it became apparent that the writer wanted to direct his/her reader(s) that the arguments would be developed on three negative impacts of globalisation on the third-world countries (economy, *culture* and *products*). Based on the scores the six assessors assigned to this stage (refer to Table 6.13 above), which ranged from 1 (one) to 2 (two), it can be assumed that they were able to identify this flaw and agreed with each other on the scores they should give to the **Thesis** stage of Essay 10a.

Another flaw that indicates lack of control of genre is found in the body paragraph of the essay. In the first **Argument** stage, for example, the *hyperTheme* or **topic** of the paragraph – *Globalisation can impact the economy of developing countries* – was not considered effective because the main idea was somewhat vague in that the statement did not clearly indicate whether the writer wanted to develop his/her

argument on the negative effects of globalisation as previously indicated in the **Thesis** stage of the essay. As a result, the writer's attempt to elaborate on the main topic of the paragraph was not successful. In the first clause of the **elaboration** phase, for example, the statement of *With globalisation the economy of the countries can be developed with cooperation with developed countries*, lacked focus, in that the statement did not support the whole purpose of the essay as stated in the **Thesis**, i.e., to argue that globalisation has negative impacts on third-world countries. Based on the scores of the six teachers allocated to this stage (see Table 6.13 above), which ranged from 1 (one) to 2 (two), it is sensible to assume that the teachers were able to identify the flaws and to reach consensus on the assigned scores.

The last evidence showing the writer's lack of control of the argument genre can be seen in the **Restatement** stage of the essay. In the **review** phase, for example, the statement of: *Globalisation can impact on developing countries in negative ways*, was sufficient and effective enough to distil the whole points of arguments as stated in the **Thesis** of the essay. However, the following clause: *The developing countries will lose their culture because of globalisation* in the **restate** phase was not effective as it did not restate the whole points of arguments previewed in the **Thesis** stage and discussed in the **Argument** stage. Based on the scores allocated to this stage of genre presented in Table 6.13 above, it can be said that the six teachers were not able to reach consensus in their marking of the Restatement of Essay 10a. The lowest assigned score was 0 (zero) and the highest was 2 (two).

From the data above (see Table 6.13), it is sensible to conclude that, overall, there was an increase associated with inter-rater reliability scores in the marking of each

stage of genre. The six teachers were able to reach consensus in the marking of the Introduction (**Issue/Thesis**), Body (**Sides/Arguments**) of Essays 8a and 10a and Conclusion (**Resolution**) of Essay 8a. In the introduction part of Essay 8a, for example, the scores the six assessors assigned to the **Issue** stage were almost identical in that the gap between the highest and lowest scores was only 0.5. Similarly, their ratings of the body paragraph (**Sides** stage) were relatively close to each other. In the *hyperThemes* or **topics**, the highest score given was 3 (three) and the lowest was 2 (two). This smaller gap indicated that the six teachers had a higher inter-assessors agreement in the marking of the *hyperThemes* of **topics** of the essay when using the SFL-informed rubric than when using the old rubric. Similarly, in the concluding paragraph (**Resolution** stage), the scores they assigned were comparatively equal. With 1 (one) score higher or lower than each other, the teachers' agreement in marking this stage of the essay was higher than that of using the existing marking rubric, as discussed in the previous section.

The similarity of the scores assigned to each stage of genre in Essays 8a and 10a indicated the existence of improved inter-assessors' consistency in essay marking using the SFL-informed rubric. However, the improvement of teachers' agreement was not evenly significant across the genre stages of Essay 10a. From the data in Table 6.13 above, it can be seen that the most improved inter-assessor agreement was in the **Thesis** and **Argument** stages. The teachers' agreement in the marking of the **Restatement** stage of the essay was lower than those of the **Thesis** and **Argument**. Perhaps, this might have been because the restatement was less well described in the intervention and the material associated in genre pedagogy. The dominance of similar scores across the genre stages of the two essays indicates the

effectiveness of the new rubric in enhancing the teachers' consistencies in essay marking.

In terms of marking the register variables of the essay, the scores the teachers assigned were compatible with each other. As seen below (refer to Table 6.14), it became apparent that the teachers' agreement was significantly high, especially with regard to the **Field**: *topic research and content* and the **Tenor**: *communication roles*, in which the score gap was not more than 1 (one) point above or below each other. This indicates that there were no serious problems found in the marking of these features of the essays. However, in the marking of the **Mode**: *written features*, the teachers' scores were not totally compatible with each other. Although the teachers were able to reach consensus in marking the written features of the high-scored essay (Essay 8a), they failed to do so when assessing the same feature in the low-scored essay (Essay 10a).

**Table 6.14 Score gap of Essays 8a and 10a based on register variables**

| Register variables | Field: topic research and content |           | Tenor: communication roles |           | Mode: written features |           |
|--------------------|-----------------------------------|-----------|----------------------------|-----------|------------------------|-----------|
|                    | Essay 8a                          | Essay 10a | Essay 8a                   | Essay 10a | Essay 8a               | Essay 10a |
| Lowest score       | 2                                 | 1         | 2                          | 1         | 2                      | 0.5       |
| Highest score      | 2.5                               | 2         | 3                          | 2         | 3                      | 2         |
| <b>Score gap</b>   | 0.5                               | 1         | 1                          | 1         | 1                      | 1.5       |

In the post-training interviews, teachers made positive comments about their essay marking capabilities. These positive comments were very largely corresponding to their improved marking consistencies. Their responses to the interview question, “*During the training, you were introduced to the Systemic Functional Linguistics, the concept of Genre, Register, Appraisal and also the SFL-informed marking*

*rubric. Now, I want to ask your opinions about the new marking rubric; how do you find the new marking rubric and the criteria?"* affirmed that the structure (*composition* in Appraisal terms) of the SFL-informed rubric related to content and choice of words had helped improve the teachers' capability. In Excerpts 45-46 below, for example, one of the teachers namely Dina, appraised the construct of the SFL-informed rubric positively in that it helped her mark the content (**Field**) of the essay:

|    |                                                                     |       |                                                                          |           |
|----|---------------------------------------------------------------------|-------|--------------------------------------------------------------------------|-----------|
| 45 | Basically, in general, the previous marking criteria we used before | don't | give me space to assess the content of the essays, the students' essays. | (Dina, 7) |
|    |                                                                     | deny  | t: -comp                                                                 |           |

|    |         |                        |                                                                                                                               |           |
|----|---------|------------------------|-------------------------------------------------------------------------------------------------------------------------------|-----------|
| 46 | But     | these marking criteria | definitely give me space to assess students' essay content you know ... not just grammatical inaccuracy and things like that. | (Dina, 8) |
|    | counter |                        | t: +comp                                                                                                                      |           |

Similarly, another teacher, Ratna, attributed the clarity of direction (*composition* in Appraisal terms) in the SFL-informed rubric to her consistency in marking student essays. She knew what each criterion meant and what was expected in each criterion (see Excerpts 47-50).

|    |           |       |       |          |                                   |            |
|----|-----------|-------|-------|----------|-----------------------------------|------------|
| 47 | I think   | it is | more  | academic | in terms of terms and also field. | (Ratna, 4) |
|    | entertain |       | force | +val     |                                   |            |

|    |                  |         |             |  |  |            |
|----|------------------|---------|-------------|--|--|------------|
| 48 | The direction is | clearer |             |  |  | (Ratna, 5) |
|    | mono             |         | force/+comp |  |  |            |

|    |                            |                              |          |  |  |            |
|----|----------------------------|------------------------------|----------|--|--|------------|
| 49 | and when we apply that ... | this is what the aspect says |          |  |  | (Ratna, 6) |
|    |                            |                              | t: +comp |  |  |            |

|    |       |            |  |                                                           |  |            |
|----|-------|------------|--|-----------------------------------------------------------|--|------------|
| 50 | It is | easier     |  | for me in marking my students' essays rather than before. |  | (Ratna, 7) |
|    | mono  | force/+cap |  |                                                           |  |            |

This clarity, as seen in Excerpts 51-53, contributes to her positive evaluation of her capabilities (Judgement/*capacity* in Appraisal terms) in assessing her students' essays.

|    |                         |                                      |                        |             |
|----|-------------------------|--------------------------------------|------------------------|-------------|
| 51 | Before, <b>probably</b> | I <b>didn't know</b>                 | what people call that, | (Ratna, 8)  |
|    | entertain               | -cap                                 |                        |             |
| 52 | <b>but</b>              | now it is                            | <b>clearer</b>         | (Ratna, 9)  |
|    | counter                 |                                      | force/+cap             |             |
| 53 | <b>I mean</b>           | <b>I have broadened my knowledge</b> | in that case.          | (Ratna, 10) |
|    |                         | +cap                                 |                        |             |
|    | pronounce               |                                      |                        |             |

In a similar vein, Whitney's evaluation of the professional learning program statement below confirmed her positive evaluation of the SFL-informed rubric as well as her capability in marking the **Tenor**: *communication roles* of students' essays. Her use of intensified positive Judgement of *capacity* (*easier*) affirmed her positive evaluation of this register variable.

|    |               |           |                                     |  |              |
|----|---------------|-----------|-------------------------------------|--|--------------|
| 54 | I, especially | also like | the Communication Roles, the Tenor, |  | (Whitney, 7) |
|    | pronounce     | +sat      |                                     |  |              |

|    |           |                      |                                                       |  |              |
|----|-----------|----------------------|-------------------------------------------------------|--|--------------|
| 55 | because   | it was easier for me | to choose the words according to their academic tone, |  | (Whitney, 8) |
|    |           | force/+cap           |                                                       |  |              |
|    | pronounce |                      |                                                       |  |              |

|    |                    |          |       |         |              |
|----|--------------------|----------|-------|---------|--------------|
| 56 | Appraisal ... yeah | that was | very  | helpful | (Whitney, 9) |
|    | concur             |          | force | +val    |              |

|    |            |                                                                                                                                                |  |  |               |
|----|------------|------------------------------------------------------------------------------------------------------------------------------------------------|--|--|---------------|
| 57 | Definitely | the Tenor, because I could choose words specifically that were not in keeping with academic tone or with the Appraisal, so the choice of words |  |  | (Whitney, 10) |
|    | pronounce  | t: +val                                                                                                                                        |  |  |               |

|    |                                                                  |  |       |         |               |
|----|------------------------------------------------------------------|--|-------|---------|---------------|
| 58 | The teaching, the seminar on identifying the choice of words was |  | very  | helpful | (Whitney, 11) |
|    | mono                                                             |  | force | +val    |               |

The teachers' opinions of the criterion related to **Tenor**: *communication roles* of the essays were, however, not equally positive. Their response to the interview question, "Talking about assessing student essays, can you tell me the feature(s) of the essays you find most difficult to mark when using this new rubric?" upheld this view. In Mirzan's statements below, for example, Appraisal was evaluated negatively as being *more difficult* to assess than the other features, but the degree of negativity was to some extent moderated or softened by his use of a low intensification *quite* and a neutral Appreciation *new*. This may suggest that the teacher needed more time and practice to build linguistic knowledge to assess the evaluative words (Appraisal features) of student essays.

|    |           |                  |            |           |                         |             |
|----|-----------|------------------|------------|-----------|-------------------------|-------------|
| 59 | I believe | the Appraisal is | a bit more | difficult | than the other features | (Mirzan, 8) |
|    | entertain |                  | force      | -val      |                         |             |

|    |                    |           |      |         |             |             |
|----|--------------------|-----------|------|---------|-------------|-------------|
| 60 | Well, because it's | quite     | new  | for me, | that's why. | (Mirzan, 9) |
|    |                    | force     | -val |         |             |             |
|    |                    | pronounce |      |         |             |             |

Similarly, in Ratna's statements below the negative evaluation made on marking Appraisal was related to her understanding of the meaning of the words both when standing alone and when related to the context.

|    |           |                                         |            |           |             |
|----|-----------|-----------------------------------------|------------|-----------|-------------|
| 61 | For me    | it's about, for one thing communication | especially | Appraisal | (Ratna, 13) |
|    | entertain |                                         | force      | -cap      |             |

|    |           |           |                   |                                |             |
|----|-----------|-----------|-------------------|--------------------------------|-------------|
| 62 | because I | need to   | look at the terms | in what way, whether it is the | (Ratna, 14) |
|    |           | entertain | t: -cap           | right one or not.              |             |

|    |                     |           |            |                         |             |
|----|---------------------|-----------|------------|-------------------------|-------------|
| 63 | And the second one, | I need to | understand | the vocabulary and the  | (Ratna, 15) |
|    |                     | entertain | t: -cap    | context of the sentence |             |

In terms of marking the structure (**Mode**) of the essay, the teachers' marking results also indicated that the teachers were, overall, compatible with each other in their judgements of this feature of students' essays. However, these judgements of the **Mode** of students' essays were not as compatible as those of the *content* (**Field**) and *communication roles* (**Tenor**). In their judgements of the Mode of Essay 8a, for example, the teachers were able to reach consensus in that the gap between the highest and lowest scores was not more than 1 (one) point. The highest score given to this feature was 3 (three) and the lowest was 2 (two). In their judgements of Essay 10a, however, the teachers did not totally agree with each other on the scores they should assign to this feature. As seen in Table 6.14 above, the discrepancy of scores was more than 1 (one) point; the highest score awarded was 2 (two) and the lowest was 0.5.

The results above were reflected in the teachers' statements regarding the marking of the *written features* (**Mode**) of the essays. Of the six teachers' accounts analysed, only one teacher, namely Dina, who perceived having a problem marking this feature of students' essays. Her statements in Excerpts 65-67, which were all realised in Judgement of *capacity* (*I find it a bit difficult* and *I am a bit confused*), reflect her opinion about her capability of assessing the textual features of students' essays.

|    |                          |                             |       |           |            |
|----|--------------------------|-----------------------------|-------|-----------|------------|
| 64 | I want to point out that | written features, I find it | a bit | difficult | (Dina, 11) |
|    | pronounce                |                             | force |           |            |

|    |     |                 |          |                       |            |
|----|-----|-----------------|----------|-----------------------|------------|
| 65 | I'm | slightly, a bit | confused | with textual cohesion | (Dina, 12) |
|    |     | force           | -cap     |                       |            |
|    |     | pronounce       |          |                       |            |

|    |                                           |       |           |             |            |
|----|-------------------------------------------|-------|-----------|-------------|------------|
| 66 | that's why the Written Features I find it | a bit | difficult | just a bit. | (Dina, 14) |
|    | pronounce                                 | force | -cap      | force       |            |

The six teachers were also able to reach consensus in their marking of the Accuracy features of the essays (see Table 6.15). The highest agreement was found in their marking of the feature related to the size of the essays, which was identified as *Adherence to Word Limit* in the rubric. Their compatibility in assessing the essays in terms of *Adherence to Word Limit* was not unforeseen. As the teachers had been advised, the student essays would receive a full score if the total number of the words did not exceed 10% below or above the word limit (300 words). As a result, all of the teachers assigned the identical scores to these essays. The teachers were also compatible in their judgements of the essays in terms of *Spelling and Punctuation* and Grammatical Accuracy, in which the score gap was not more than one point apart, suggesting that teachers did not have problems in assessing these features of students' essays.

**Table 6.15 Score gap of Essay 8a and Essay 10a based on accuracy of the essay**

| Accuracy         | Spelling &Punctuation |           | Grammatical Accuracy |           | Adherence to word limit requirement |           |
|------------------|-----------------------|-----------|----------------------|-----------|-------------------------------------|-----------|
|                  | Essay 8a              | Essay 10a | Essay 8a             | Essay 10a | Essay 8a                            | Essay 10a |
| Lowest score     | 2                     | 1         | 2                    | 1         | 3                                   | 3         |
| Highest score    | 3                     | 2         | 3                    | 1         | 3                                   | 3         |
| <b>Score gap</b> | 1                     | 1         | 1                    | 0         | 0                                   | 0         |

Interestingly, the amount of feedback varied with essay scores. In general, the teachers gave more feedback on the low-scored essay (Essay 10a) than the high-scored one (Essay 8a). In the low-scored essay, for example, most of the feedback was on the punctuation such as commas (,), full stops (.) and grammatical mistakes such as passive voice, singular/plural noun and S-V agreement. These are the

criteria that are the most obvious and familiar, so therefore are the ones with most agreement. In the high-scored essay, however, there were few corrections made regarding the two features of the essay. The only feedback on grammar was found in the **elaboration3** of **Side2** in which the sentence: *Another negative impact on local cultures in developing countries which are affected by foreign cultures through social media* was changed into *Another negative impact is that local cultures in developing countries are affected by foreign cultures through social media*.

To conclude, teachers' assessment of students' essays using the SFL-informed rubric has been shown to shift quite significantly. This was evident in the increased agreed-scores among the six teachers, from 25.6% when using the existing marking criteria to 39.3% when using the SFL-informed criteria. Detailed analysis of the scores teachers assigned to each criterion at the level of genre, register and expression has also shown that the teachers, had shared an understanding of the criteria. Compared with the results of their perceptions, which revealed positive judgements of their capacities, it can be concluded that teachers' actual marking practices reflected their perceived capabilities.

## 6.3 Summary

The purpose of this chapter was to examine teachers' opinions and experiences of using two different marking rubrics: the existing marking rubric provided by the language institution where the teachers' worked (pre-intervention), and the SFL-informed rubric (post-intervention). The findings can be summarised into two themes. Firstly, teachers were well aware of their marking capabilities when using

the two rubrics. Their comments about their capabilities were predominantly negative when using the existing marking criteria. These negative opinions of capacity were reflected in the low inter-assessors agreement counts in their actual pre-intervention marking activities. In contrast, after the training, teachers' perceived their marking consistencies using the SFL-informed rubric very positively, and these positive evaluations were reflected in the high inter-assessors agreement counts. As discussed above, the most improved inter-assessors agreement was found in the **Thesis** and **Argument** stages (genre), topic research and content (**Field**), communication roles/Appraisal (**Tenor**), grammatical accuracy and spelling. However, teachers tended to have less agreement in their marking of the **Restatement** stage (genre) and written features (**Mode**) of students' essays. Secondly, teachers were able to recognise the gap between marking, teaching practices and students' needs. Teachers' recognition of the shortcomings of the existing rubric in terms of choices of criteria and what the students were capable or not of doing also helped the design of the SFL-informed rubric. The findings and their implications are discussed in the following final chapter.

## **Chapter Seven**

### **Discussion and conclusion**

The previous chapter has addressed all the research questions driving this thesis. It has clearly shown the criteria and features of students' essays that teachers identified as most problematic to assess and the teachers' perspectives of the two marking rubrics in relation to their marking consistency. As stated earlier, the study set out to investigate teachers' marking consistency and enhance their assessment practices surrounding academic essays in an Indonesian language school. This was done by developing and introducing more explicit and linguistically informed marking criteria through an intervention. An action-oriented case study was adopted to capture teachers' marking experiences and the intervention. The study comprised two sets of data: argument essays written by 20 English for Academic Purposes (EAP) students and interviews with teachers about their marking practices. The thesis drew upon Systemic Functional Linguistics (SFL) theory (see, for example, Halliday 1978; 1985a, Halliday and Hasan 1985, Halliday and Matthiessen 2004, Martin 1984; 1997) to inform the design of the new rubric and the intervention, and to analyse students' texts and teachers' accounts. The purpose of the present chapter is to discuss the key findings in the light of the given theory and relevant literature. The final section of the chapter draws some relevant conclusions, implications and the directions for future research.

This section has discussed the findings related to the research questions posed in Chapter One. Teachers' marking reliabilities, when using both the existing and the newly designed rubrics have been discussed. The results of the study have shown

teachers' improved consistencies when using the new rubric, both in terms of the overall scores assigned to students' essays and on criteria level (such as genre stages, register and accuracy). The study examined teachers' opinions about the marking rubrics used, the intervention and their consistencies before and after the training. In line with their improved reliability, teachers' perceptions of the SFL-informed rubric were predominantly more positive than those of the old rubric. This is shown by the prevalence of positive evaluative comments such as "It is clear", "it is more complete", "This is more comprehensive" (positive *composition*), "The marking rubric is much more useful" and "I find it quite helpful" (positive *valuation*) in the teachers' talk about the new rubric.

To summarise this section, it can be said that the results of the study are quite encouraging. First, most of the raters expressed positive attitudes towards the SFL-informed rubric and regarded it as a useful component of the training program. Secondly, most of the teachers agreed that the intervention has helped them improve their essay marking reliability, and at the same time, encouraged them to think more carefully about what they were doing in their marking practices.

### **7.1. The relationship between teachers' perceptions and their assessment practices**

The research literature on assessment has noted the importance of analysing teachers' perceptions. Some focuses on teachers' opinions in relation to their preparedness and support (e.g., Brannan and Bleistien 2002, Faez and Valeo 2012) and students' and teachers' needs (I. Lee 2008b; 2010, Matsuda, Saenkhum et al. 2013, Nakamaru 2010). Others discuss teachers' perceptions about their teaching

philosophies and feedback (e.g., A. Cohen and Cavalcanti 1990, Ferris, Brown et al. 2011, Hyland 2013, I. Lee 2008a; 2009b, Montgomery and Baker 2007). However, as we have seen in the literature, very little research has been conducted on the question of the relationship between teachers' perceptions and their assessment practices in the field of academic writing.

In this section, the main research findings are drawn together in a discussion that aims to contribute to Systemic Functional Linguistics theory and the current assessment policy and practice in EAP programs in Indonesia. The present study has argued that it is important to examine the perceptions of those involved in the assessment process regarding the marking rubric used and of their marking capabilities. Because previous research has noted the importance of focusing on teachers' perspectives, the present study helps illuminate these perspectives by incorporating the teachers' perceived capabilities and realities in their marking practices to the design of new marking criteria.

### **7.1.1 The influences of the linguistically informed assessment criteria on inter-rater reliability**

The literature has underlined the importance of rater reliability in assessment, in that the reliability of assessment is determined by the consistency of the assigned scores across different assessors and occasions (Moskal and Leydens 2000). As reliability is prone to raters' interpretations of the marking criteria (Barkaoui 2007; 2010b), the literature has further suggested that the key solution to the improved reliability of assessment is "the provision of clear marking criteria and schemes" (Gipps 2012, p.

137). The present study has argued that a linguistically informed assessment rubric can improve teachers' marking consistency.

The results of the study, which examined the essay marking practices of six Indonesian EAP teachers in terms of inter-rater reliability, have indicated a strong relationship between the marking rubrics used and the level of the agreed scores among the teachers. In the two marking sessions – pre-intervention using the existing marking criteria, and post-intervention using the SFL-informed criteria – the teachers' consistency when using the SFL-informed rubric outperformed consistency when using the existing rubric. This improved consistency, as has been summarised in the previous chapter (see Figure 6.3), occurred in terms of both the teachers' individual and collective performance. After the intervention, the teachers were, in general, able to reach consensus on the scores of four out of ten essays, while before the training no more than two essays received similar scores.

The SFL-informed criteria were also considered more explicit and easier to understand. The findings, which have been summarised in Figure 6.3 of the thesis, show that the reliability rate was more evenly distributed across the six teachers in that the gap between teachers' rating consistency was smaller than that when using the existing marking criteria. The discrepancy of teachers' performance when using the old marking rubric, indicated by their consensus estimate of inter-rater reliability rate (Stemler 2004), was 22%. In contrast, this discrepancy became narrower (14%) when using the SFL-informed rubric. This means that the teachers' marking capabilities became more evenly distributed when using the new rubric and became more aligned with one another. This greater level of consistency indicates that the

teachers developed similar perspectives on most of the criteria used in the SFL-informed rubric. At the same time, the findings also affirm what the thesis has argued – that is, the explicit and linguistically informed assessment criteria, which became the shared knowledge of the teachers, would improve consistency of marking among assessors.

### **7.1.2 The importance of understanding teachers' perspectives**

Teachers' opinions, according to the research literature (e.g., Barkaoui 2007, Crocker 2002, Norris 2008, Ryan 2002, P. Winke 2011), should be heard and not ignored. By taking teachers' perceptions into consideration, "standardized testing programs can improve their ESL or EFL tests and better ensure they are broadly valid – that the tests produce reliable scores and have positive impacts" (P. Winke 2011, p. 653). In educational assessment, teachers are important stakeholders who have responsibilities for the qualities and reputations of a system in an educational institution.

Teachers' reliability in assessing their students' performance on tests or tasks such as essay writing is crucial for two main reasons. Firstly, the unreliability of assigned scores, which may result from the scoring items or criteria and teachers' different understanding of the criteria, can misinform students' capabilities for given tasks, and may eventually 'stain' the reputation of the schools or even the career of the teachers (P. Winke 2011). The findings of the present study show that some of the teachers expressed their concerns about their lack of understanding of some of the criteria in the existing rubric and the unavailability of score traits that inform the characteristic of each score. For example, Dina stated her uncertainty about the

meaning of *Readability* in the original rubric and the ways to assess it in students' essays. Another example is Ratna's statement of "It **is** just number", indicating her disappointment with the rubric due to the absence of score traits to guide her assessment of students' writing. Secondly, and probably the most important reason of all is that teachers have unique insight into the collateral effects of the marking criteria used on their students' performance on the tasks given. They administer the teaching and assignments, therefore they know the students' needs in terms of strengths and weaknesses (Matsuda, Saenkhum et al. 2013).

The findings of the present study show that teachers are well positioned to recognise discrepancies between students' tasks and the tool (rubric) used to assess the tasks. One example is Mirzan, who puzzled over the absence of a criterion that focuses on 'the overall structure' of student essays despite the fact that he and his colleagues always emphasised the importance 'the overall structure' in their teaching of EAP classes. This indicates the discrepancy between pedagogy and assessment. Thus, it is sensible to say that the teachers are well informed on whether or not the writing construct is adequately represented by the scale used (Knoch 2009), and they certainly have reasonable insights into the marking practices and the rubric.

The present study has shown that teachers' perceptions are valuable pieces of information to inform the design of assessment criteria. Together with the more explicit criteria informed by the SFL, the accommodation of teachers' insights into their problems in assessing students' essays and students' needs had succeeded in enhancing teachers' quality of marking. This is indicated by the improved rate of their inter-rater reliability from an average of 25.5% to 39% (see Chapter Six for the

detailed information of teachers' reliability). As Moskal and Leydens (2000) argue, improved score consistency among teachers can be seen as equal to improved reliability.

In addition, the accommodation of teachers' perspectives in terms of marking practices, teaching practices and students' capabilities to the design of SFL-informed criteria has, at the same time, established the validity of the assessment – that is, the criteria measured what they were supposed to measure and were justified in terms of their outcomes (Bachman 1990, Hughes 2003). Before the intervention, most of the teachers felt dissatisfied with the criteria used in the rubric, and hoped that these criteria would change to better reflect what was actually taught or assessed in the students' essays. As suggested by Knoch (2009), the inadequacy of criteria that represent the features of students' writing in a given genre can be argued to have reduced the construct validity of the rubric used. The teachers' dissatisfactions were shown by the predominance of negative opinions about the existing marking criteria identified as negative *composition* in Appraisal analysis. In the present study, for example, teachers' comments reflecting their dissatisfaction with the compositional aspects of the rubric can be categorised into four main stands. The first is concerned with the selection of criteria (e.g., the absence of a content-related criterion, *Subject-Verb Agreement*, the conflation of *Vocabulary* with *Spelling and Punctuation*). The second is related to the relative weighting of the criteria (e.g., the maximum score for *Subject-Verb Agreement* was the same as that of *Vocabulary*, *Spelling and Punctuation* put together). The third is the clarity of the criteria such as *Readability* to teachers. The last one is the transparency of score traits, i.e., there is no descriptor that informs the specific quality of each score.

As *composition* is concerned with perceptions of order or balance (Martin 2000), this finding can be interpreted as a sign of the participants' dissatisfaction regarding the selection of criteria in the existing marking rubric and the way they are arranged and valued. In other words, as Knoch (2008) argued, the criteria in the existing marking rubric did not reflect what should be assessed in the students' essays. In contrast, teachers' prevalent positive opinions of the construct of the SFL-informed rubric indicate their satisfaction with the way the criteria were selected, weighted and arranged. The detailed comparison of teachers' positive and negative opinions of the two rubrics is summarised in Figure 7.1 below:

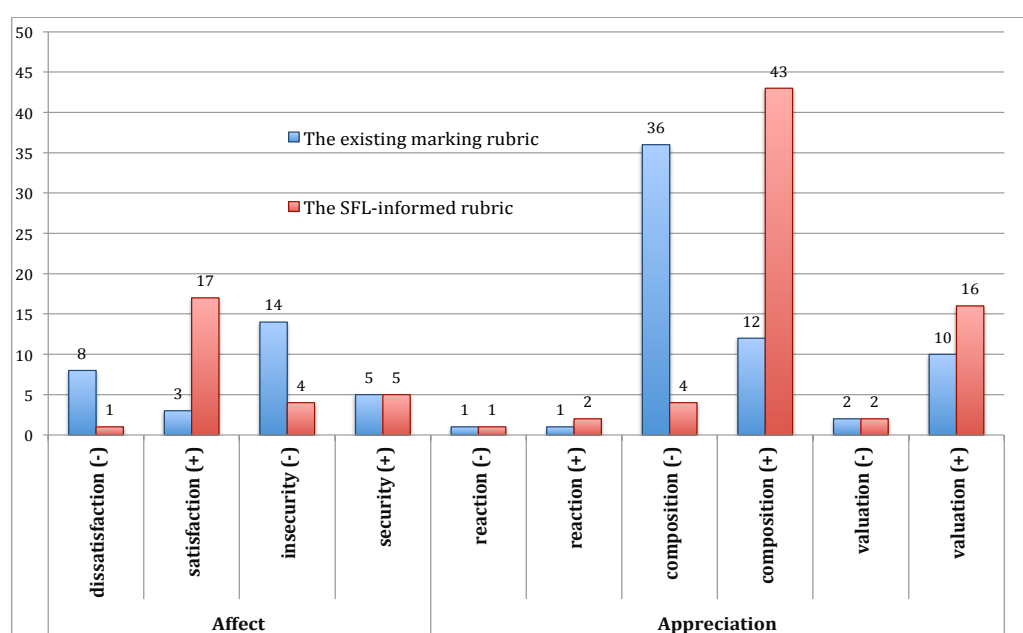


Figure 7.1 A summary of teachers' perceptions of the two rubrics

Despite the teachers' negative feelings towards the compositional aspects of the existing marking rubric, they appreciated it very highly. This was shown by the predominant choice of positive *valuation* such as *helpful* or tokenised *help*, over the negative ones. A possible explanation for this is that the existing assessment rubric provided a benchmark for their marking practices. The statement made by one of the

teachers, Mirzan, can be taken as a rationalisation of the positive evaluation given by the teachers. His statement that, “I believe a lot of teachers would be more subjective without the help of the rubric” expressed his perception of its usefulness. Simply put, the statement above suggests that having a rubric, irrespective of its imperfections, was better than not having one at all.

### **7.1.3 Teachers’ opinions and their marking practices**

The first and second research questions concern the most difficult or problematic aspects of students’ essays the teachers identified when using the existing marking rubric. These perceived difficulties are, overall, reflected in the teachers’ actual marking practices (shown by their inter-rater reliability counts), and at the same time answered the first two research questions posed earlier in Chapter One of the thesis. On the question of the features with the least inter-assessor consistency, this study found two main sources of inconsistency in marking students’ texts. The first source was found in a paragraph level of structure in the body paragraphs of the students’ essays, and the second one was in the vocabulary used at the discourse level. In the body paragraph, for example, the discrepancy of scores among the teachers was found in the *Topic Sentence* and *Supporting Sentence* of the essays. This finding was in line with the results of teachers’ perceived difficulties in marking. One of the reasons for this inconsistency was caused by the separation of *Topic Sentence* and *Supporting Sentence*, which the teachers perceived as two related elements. This proposition was strongly affirmed by one of the teachers, Dina, who perceived that *Topic Sentence* and *Supporting Sentence* are interrelated, and should not be marked separately. In terms of the concluding stage of the essays, however, teachers’ perceptions of their capabilities were not reflected in their marking performance.

Although none of the teachers perceived it difficult to mark the *Conclusion* stage, the results showed that they were unable to reach consensus in the marking of this section of students' essays.

The present study has shown that teachers' opinions and their actual consistencies are not necessarily contradictory. This finding affirms the results of some earlier studies by Cohen and Calvacanti (1990) that also found a strong link between teachers' self-assessment and their actual performance. Similarly, Hyland's (2013) study shows a strong relationship between teachers' perceptions and their actual feedback on students' essays. The strong correlation between the teachers' perceptions and the results of their inter-rater reliability counts establishes the construct validity of the rating scales. Previous studies have proposed the design of rating scales based on empirical investigation of actual writing samples (Arkoudis and O'Loughlin 2004, North and Schneider 1998, Turner and Upshur 2002). As explicated in Chapter Five, the selection of criteria in the new rubric was informed by Systemic Functional Linguistics, with a focus on text-context relations, as explained in Chapter Three. Thus, the new rubric assessed the students' essays more systematically in the light of genre (e.g., purpose and stages developed to achieve the purpose), register variables of *field* (e.g., the topic/issue and content), *tenor* (e.g., formality and choice of evaluative language to encode critical tone) and *mode* (e.g., Theme, conjunctions and references to create cohesion and coherence), and also *accuracy*, which comprised general academic and mechanical features of the essays (e.g., tense, spelling, punctuation and word limit). Empirically, the design of the new rubric was informed by analyses comparing the teachers' perceived difficulties in

assessing students' essays to their marking practices (RQ1 and RQ2) prior to using the SFL-informed rubric.

In this study, comparing the features of the essays that teachers found most difficult to mark with the actual scores given to the features also showed *vocabulary* as another source of teachers' essay marking inconsistency. It was also perceived by the teachers as a problematic area. These perceptions were reflected in the results of their marking practices on this particular feature of student essays. The agreement of perceptions and practices is an indication of the teachers' awareness of problems in assessing students' vocabulary, particularly in terms of range and accuracy of vocabulary usage. As Ruegg, Fritz and Holland (2011) argue, although vocabulary in EFL/ESL writing is often assessed, it is yet unclear how well raters are able to assess the range and accuracy of vocabulary usage (p. 64). Even if we assume that teachers have high proficiency in using and marking students' vocabulary, the compatibility of scores to each other remains refutable due to the fact that it is combined with *spelling* and *punctuation* in the existing marking rubric.

That marking vocabulary is deemed difficult and proven to be inconsistently assessed may not be attributed to the fact that five out of six assessors are not native English speakers. When comparing the scores given by the five non-native English speaking (NNS) teachers, the gap between the highest and the lowest scores was 1.5 points, which was considered as having no inter-rater reliability, the difference being more than 1 (one) point above or below (Stemler 2004). This gap was even narrower when compared to the native English-speaking (NES) teacher, Whitney. Thus, it can be argued that, in terms of scores assigned, the similarity in language background

does not always equal similarity in teachers' perceptions of the appropriateness of the vocabulary used, particularly among NNS teachers who are fluent in written and spoken English. Perhaps, the score variability was more related to the teachers' differed understanding of the given criterion (Ruegg, Fritz et al. 2011). Then, it can be argued that the incompatibility of scores assigned by the six teachers to this criterion (vocabulary) was not influenced by the teachers' language backgrounds (Indonesian-English). This finding, to some extent, confirms Johnson and Lim's (2009) finding, that language background-related bias was not found in teachers' ratings of students' vocabulary.

Based on the evidence gathered from the teachers' talk, it can be suggested that the problems in marking vocabulary are more related to the difficulty experienced by the teachers to separate meanings related to Field from those related to Tenor. As all the teachers in this study had taught the subject for at least three years, it can be assumed that they had adequate knowledge to mark the content-related vocabulary (i.e., lexical words related to globalisation) in the students' essays. Research in vocabulary and writing quality has indicated the highest correlation of error-free lexical variation with the scores assessors assigned (Engber 1995). This means that lexical choice and its correct usage significantly affect the teacher's judgement of an essay – that is, the use of wrong words (lexical errors) tends to be marked harshly or negatively by assessors (Santos 1988). In this study, however, it is yet uncertain whether the vocabulary the teachers assessed also included the evaluative words, and if this is the case, the scores assigned to these two remained unclear as it was mixed with *Spelling* and *Punctuation* as one criterion in the existing marking rubric (see Table 6.6).

When thinking about what constitutes quality writing, vocabulary is of special importance for second-language writers, because this may be the crucial factor in determining the reader's ability to understand and evaluate the writing (Engber 1995). As was proposed earlier in this study, teachers' recognition of *tenor* related vocabulary combined with their expertise in the *field*-related vocabulary would help their marking lexical choice of students' essays. In the intervention, the teachers were trained to identify both the *field*- and the *tenor*-related vocabulary. In a clause "One of the negative impacts of globalisation is tax evasion", for example, teachers' recognition of two nominalised noun groups 'one of the negative impacts of globalisation' and 'tax evasion' connected by a relational process/verb 'is', helped them identify what the clause is about (e.g., the relationship between globalisation and tax evasion). At the same time, within the same clause, the teachers could recognise the student's negative evaluation of the topic being discussed (negative impacts of globalisation). The results of the post-intervention reliability counts affirm this proposition. As explicated in the previous chapter, the teachers were able to reach consensus in their marking of *field*- and *tenor*-related vocabulary, in that the score gap was not more than 1 (one) point apart.

The findings of the present study have indicated that teachers' talk about their assessment performance when using the existing and the SFL-informed rubrics (pre-/post-intervention) was reflected in the inter-rater reliability counts. The correspondence between their opinions and performance can be seen both in terms of the overall scores of the essays and in the assigned scores for each criterion. When using the existing rubric, for example, most of the teachers' comments about their reliability in assessing students' essays were negative. These were shown by

the prevalence of negative Judgement of *capacity* (see Chapter Six) when evaluating their marking capabilities. In contrast, when using the SFL-informed rubric, teachers' accounts of their marking reliability were more prevalently positive, shown by the dominance of comments identified as positive Judgement of *capacity*. This positive evaluation suggests their increased confidence, and at the same time, points to the clarity of the new marking rubric.

#### **7.1.4 Teachers' perspectives and marking ethics**

In addition to teachers' perceptions of their skills and confidence, the thesis has been able to shed light on teacher agency in terms of their commitment to marking and loyalty to the institution. Duranti (1997) argues that evaluation is important in the construction of agency, and it must be connected to both performance and morality. Analysis of teachers' evaluations of their assessment practices in terms of performance (*capacity*), as discussed previously, has affirmed the significance of incorporating their perspectives to the design of marking criteria. As we have seen earlier, teachers' perceived difficulties in marking the vocabulary and content of student essays have been answered in the SFL-informed rubric under the criterion of *field* and *tenor*. Similarly, their complaints about the separation of *topic* and *supporting sentences* in the existing rubric were addressed by including them in the criteria focused on the rhetorical function of the essay under the criterion of *genre* in the SFL-informed rubric. Similarly, in terms of morality or ethics, teachers' endeavours to remain virtuous can be discerned as their loyalty to the institution, i.e., they did not want to be critical of the existing marking policy. As the Appraisal analysis shows (see Chapter Six), teachers' talk about the ethical aspects of their marking practices was evident when they explained their choices with reference to

institutional policy and practice such as “I do mark according to these criteria”, “I try my best to stick to the rules” and “I have to do that based on the rubric” (Judgement of *propriety*). This indicates teachers’ strong awareness of their routine rating behaviour.

Agency can be seen as the actual performance of the routine by specific people, in specific times and places (Feldman and Pentland 2003). Teachers create and enact routines. While these routines helped to frame and sustain their marking practices, the teachers involved in the routines did so with the possibility of individual choice. In other words, they could choose, at least discreetly, to carry out the aspects of their marking routine in the ways they saw fit, based on their own goals, expertise and capacity. However, as the findings of the study have shown, the teachers were loyal to the principal values of the marking policy set by the institution they worked in. Despite that, there are times when the institutional assessment policy, which in this case represented by the existing marking rubric, clashed with the teacher’s own value, particularly when they perceived mismatches between their teaching practices with the assessment rubric (Davison 2004). Examples are Mirzan and Jelita, who claimed the importance of assessing the content of students’ essays despite their awareness of the absence of such a criterion in the existing rubric. This means that marking routines can be perceived as the teachers’ way of responding to the present circumstances and reflecting the meaning of actions for future realities (Emirbayers and Mische 1998). In other words, the teachers’ decisions to include *content* as a criterion can be seen assumed as rooted from their expectation of an ‘ideal’ marking rubric.

### 7.1.5 Improved teachers' linguistic knowledge

SFL-informed criteria had successfully provided teachers with knowledge about language (KAL) and tangible textual evidence that provided justification for their marking of students' essays. They also enabled teachers to make links between pedagogy and assessment very explicitly. The teachers' concerns with the unavailability of a content-focused criterion in the existing rubric were, for example, answered in the SFL-informed rubric under the criterion known as the **field** of the essay. During the intervention, the teachers were trained to identify this feature of their students' essays by analysing elements of Transitivity such as the participants, which are typically realised by noun groups and/or nominalisation in the texts. In academic writing such as argument essays, students are expected to present specialised content that is technical, dense and abstract (see, for example, Fang and Wang 2011, Halliday 1985a, Martin 1985, Ravelli 1988, Ventola 1996). The teachers' examination of the noun groups in students' essays in terms of technicality, abstraction and expanded/unpacked nouns helped them identify the content of the texts. Thus, it is not surprising that the teachers appraised the SFL-informed rubric positively because it offered a 'focus on the content' or 'provide space for the content'. Similarly, teachers' knowledge of the **tenor** helped them analyse voice in student texts; whether the topic of argument (the effects of globalisation in developing countries) was evaluated positively or negatively, and how the voice was used to interact and convince the reader(s). Finally, the teachers' knowledge about the **mode** of the text assisted with the assessment of the quality of student essays in terms of organisation and development of ideas. For example, different kinds of Themes indicate different approaches to text organisation.

Teachers' examination of Themes and their progressions provided perceptible evidence of the flow of ideas in the arguments a student made and the quality of argument; whether the student simply reproduced his/her knowledge about the topic (globalisation) by using mostly Theme reiteration or whether he/she combined it in zigzag and multiple Rheme patterns (Eggins 2004, Ravelli 2005) to show evidence of rhetorical crafting, which Bereiter and Scardamalia (1987) called 'knowledge transforming'.

To sum up, the present study has succeeded in incorporating information gained from teachers' comments about their marking experiences into the design of the new assessment rubric. Duranti (2006) argues, "any act of speaking [and writing] involves some kind of agency or perspective, regardless of the speaker's intentions and the hearer's interest or collaboration" (p. 451). In the present study, teachers' comments have clearly shown their agency in terms of the rubric used, reliability and difficulties in marking. In the data, teachers displayed their attitudes, feelings, and wishes about their assessment practices, which undoubtedly affected the design of the more informed criteria that represent their teaching of academic writing to students. Thus, teacher agency is considered as an important part of the rubric design along with the underlying systemic linguistic theory, which informed the selection of criteria used in the intervention. The use of specimen papers taken from the actual work of EAP students enrolled in the language course has resulted in the teachers' familiarity with the sample texts used in the process of identifying the criteria of quality essays. The training program focused on teachers' shared knowledge of the SFL-informed criteria, i.e., what each criterion means, and also teachers' identification of the criteria in the sample essays used. Thus, it is sensible

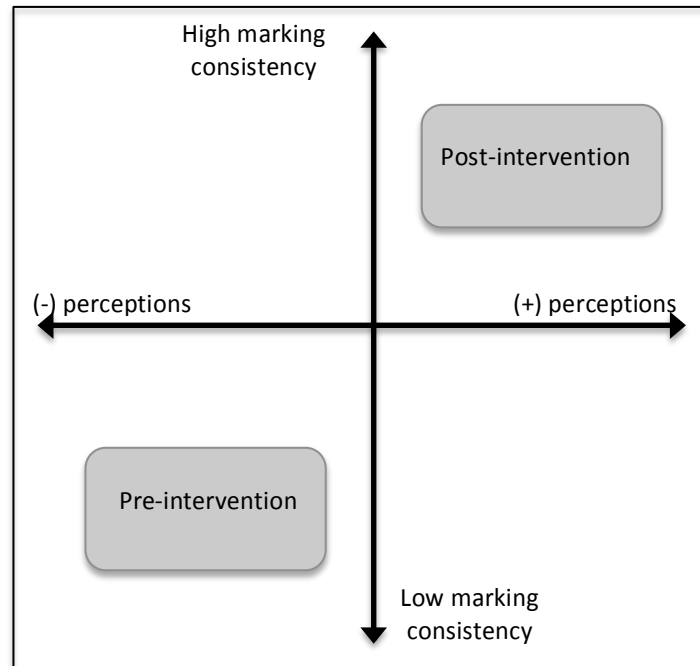
to say that the teachers were involved in the design of the new rubric and were very receptive because the idea was not imposed on them from the outside.

As the results of the study revealed, the teachers, overall, appreciated the SFL-informed rubric far more positively than the old one. These positive appreciations resulted from their satisfaction with the selection and proportion of criteria in the new rubric, among which were positively evaluated as ‘ideal’, ‘detailed’, ‘helpful’ and ‘clear’. The Appraisal analysis of the teachers’ accounts has affirmed these positive evaluations and shown that the teachers’ positive perceptions of the new rubric were predominantly focused on the structures and the criteria of the rubric, which were detailed and carefully arranged. In this study, the teachers’ preference for the more detailed rating scale affirms Knoch’s (2009) study in that they knew what each criterion meant and what was expected in each criterion. Despite the teachers’ claims on the clarity of the criteria, their inability to have significantly higher agreement rates shows that the inter-rater reliability is not simply determined by the constructs of the rubric alone. There seems to be other factors influencing the teachers’ decision-making. Perhaps, the teachers also turn to their own implicit knowledge and beliefs, especially when their personal assessment and standardised criteria are in conflict (Davison 2004).

The results of the study suggest that teachers are better placed to identify the features of a successful essay. This was reflected by the improved reliability of scores among the six teachers. In the following section, I will summarise some characteristics of quality essays, which were identified in the new marking rubric.

To reiterate from the findings in the previous chapter, successful essays have several shared characteristics. Firstly, they have good organization of stages and phases to achieve the purpose of essay writing. This is identified as genre in the new rubric. Secondly, they show mastery or knowledge of the topic being discussed (Field), evaluate the phenomena or issues using less personal language (Tenor), are lexically dense and have simple yet effective sentence structures, characterized by relational processes connecting two nominal groups, and they are cohesive and have clear progression of ideas (Mode). Field, tenor and mode are identified as the register variables in the new rubric. Finally, quality essays are characterised by correct use of grammatical conventions, identified in the new marking criteria as accuracy.

Returning to the overall purpose of the study, the following conclusion can be drawn. Bachman and Palmer (1996) suggest that “the most important consideration in designing and developing a language assessment is the use for which it was intended” (p. 17). We need, therefore, to remember that the purpose of this study was to explore the use of a linguistically informed rubric and professional learning program to enhance consistency among assessors. Most evidence speaks in favour of the project, both in terms of the teachers’ consistency, which was examined based on their consensus estimate of inter-rater reliability, as well as in terms of construct validity of the scale used. As illustrated in Figure 7.2, teachers’ negative perceptions of capability when using the existing marking rubric are reflected in the low consistency rate among the six teachers. Similarly, the prevalence of positive opinions of capacity of assessing students’ essays by using the SFL-informed rubric is translated into a higher consistency rate.



**Figure 7.2 Diagrammatic illustration of comparative shift in teachers' marking practices**

As shown in Figure 7.2, the teachers in this study had all moved from the low-marking consistency end towards the high-marking reliability on the vertical axis, with some teachers slightly further along than others. Jelita and Yusuf, for example, were at the lowest end in their inter-rater reliability counts during pre-intervention marking practice. However, after the intervention, they managed to catch up with, and even surpass, their colleagues' levels of consistency. As a result, the reliability rates were more evenly distributed across the six assessors. Their perceived accomplishment and confidence had also shifted from frequent negative perceptions to predominantly positive opinions. It is, therefore, sensible to conclude that the intervention had significant impact on rater consistency and had successfully accomplished its purpose.

### 7.1.6 Summary

This section has discussed the findings related to the research questions posed in Chapter One. Teachers' marking reliabilities, when using both the existing and the newly designed rubrics have been discussed. The results of the study have shown teachers' improved consistencies when using the new rubric, both in terms of the overall scores assigned to students' essays and on criteria level (such as genre stages, register and accuracy). The study examined teachers' opinions about the marking rubrics used, the intervention and their consistencies before and after the training. In line with their improved reliability, teachers' perceptions of the SFL-informed rubric were predominantly more positive than those of the old rubric. This is shown by the prevalence of positive evaluative comments such as "It is clear", "it is more complete", "This is more comprehensive" (positive *composition*), "The marking rubric is much more useful" and "I find it quite helpful" (positive *valuation*) in the teachers' talk about the new rubric.

To summarise this section, it can be said that the results of the study are quite encouraging. First, most of the raters expressed positive attitudes towards the SFL-informed rubric and regarded it as a useful component of the training program. Secondly, most of the teachers agreed that the intervention has helped them improve their essay marking reliability, and at the same time, encouraged them to think more carefully about what they were doing in their marking practices.

## 7.2 Conclusions and implications

The previous section has discussed the findings and argued that the research study, which set out to investigate and enhance assessment practices surrounding academic essays in an Indonesian language school, has been able to achieve its purpose. As explicated above, the enhancement of the quality of EAP teachers' assessment was done by developing and introducing more explicit and linguistically informed marking criteria through an intervention. Drawing upon the tools of analysis provided by SFL theory e.g., Clause complex, Lexical density/Nominalisation, Theme and Appraisal system (e.g., Halliday 1978, Halliday 1994, Halliday and Hasan 1985, Martin 1984) to inform the design of the new rubric and the intervention, and to analyse students' texts and teachers' accounts, the present study has succeeded in identifying teachers' marking problems by incorporating their perspectives into the design of the new assessment rubric and conducting a rater-training program.

The findings were reviewed and discussed in the previous chapter. The thesis has argued that the utilisation of a marking rubric based on the coherent model of language offered by SFL would lead to a better assessment practice. The present study demonstrates that there was quite a significantly improved consistency among the six teachers when using the SFL-informed rubric, from an average of 25.6% (when using the existing marking rubric) to 39% when using the new rubric. This improved reliability was in line with the teachers' accounts of their marking practices using the new rubric. The teachers' evaluations of the SFL-informed rubric and of their marking capabilities, which were analysed using the Appraisal system

(e.g., Martin and Rose 2007, Martin and White 2005, White 2012), were prevalently coded as positive Appreciation and Judgement respectively. This concluding section discusses implications for the assessment of academic writing. It then summarises the contribution that the present thesis makes to the area of assessment, particularly in EAP programs in an Indonesian context. It then ends with a discussion outlining the limitations of the study and suggesting possibilities for further research.

### **7.2.1 Implications for assessment practice**

What emerges from the study is that the utilisations of linguistically informed assessment criteria that become shared knowledge of teachers are required in improving assessment quality. Teachers' enhanced linguistic knowledge can lead to a better understanding of each criterion used in the assessment rubric, and this understanding will eventually lead to reliable and consistent judgements of the students' writing. Fang and Wang (2011) argue that teachers need to focus on equipping themselves with deep knowledge about how language works in different genres and registers, and then use that knowledge to guide them in identifying systematic linguistic evidence that supports their assessment of students' texts. This shared knowledge will also be beneficial to planning subsequent teaching and remediation that addresses students' needs when writing in a given genre.

In terms of communicative quality in students' texts, the findings of the present study suggest that the evaluative language in essays plays a central role in assessing the quality of writing, a feature which is well documented in studies of academic writing (e.g., Hood 2004; 2010, S. H. Lee 2006; 2007, Nakamura 2009). The study shows that quality academic essays feature particular evaluative language

preferences such as appreciation of things and ideas rather than people and their behaviour, more ‘objective’ evaluations rather than overtly subjective personal feelings and opinions. In this respect, learning and teaching practices focusing on these linguistic preferences would effectively prepare students for the writing tasks. For this purpose, teachers’ knowledge of these features is seen not only as critical to marker consistency but also as an important aspect in providing instruction that improves students’ writing skills and promoting their academic success.

With respect to assessment practices, particularly in an Indonesian educational context, there is a need to shift the focus of attention from such issues as matters of accuracy (*spelling, punctuation, capitalisation, subject-verb agreement and verb tense*) to more systematic and meaning-oriented elements – *purpose, content, interaction* and *organisation* – of the students’ essays. Some previous studies in Asian educational contexts such as Indonesian (Emilia 2005, Sutojo 1994), Thai (Kongpetch 2003, Payaprom 2012) and Vietnamese educational contexts (Dang 2002) have demonstrated the effectiveness of SFL-informed genre approaches to enhance students’ academic writing ability across different genres (e.g., Exposition, Discussion and Report). When students are taught and required to write in a particular genre, it is therefore sensible to ensure that their writings are assessed more systematically based on the purpose and corresponding linguistic features informed by the SFL genre theory. Close alignment between instruction and assessment will facilitate better learning outcomes. The study has shown this is something achievable when applied in a small-scale educational context, such as a language institution where the study was conducted, or at tertiary level of education where the lecturer is responsible and has more autonomy in the assessment of the

course. In the context of primary and secondary education, the application of this assessment practice is more complex as the teaching-learning processes are highly dictated by the curriculum prescribed by the *Depdiknas* (the Indonesian Department of Education). Thus, education policy in terms of curriculum design needs to be refined in order to bring about the alignment. For teacher education, for example, pre-service teachers need to be pre-equipped with knowledge of systemic language features so as to prepare them to teach and assess their students' writings more systematically, and, at the same time, provide instruction that enhances students' writing achievement. Similarly, for in-service teachers, workshops that focus on enabling teachers' assessment of students' writing based on tangible textual evidence need to be conducted.

### **7.2.2 The contributions of this thesis**

The present study has stimulated a discussion on the correlation between teachers' perceptions and their marking practices. This study has managed to add to the relatively small number which have done so (e.g., Eckes 2008; 2012, Hyland 2013, I. Lee 2009b, Pecorari and Shaw 2012) by providing a case study of an intervention using a linguistically informed rubric and rater training program. It also has examined the nature of the markers' experience of the intervention as well as its impact on their practices. Furthermore, this study has provided insights into the extent to which SFL-informed marking criteria can improve quality of assessment, which was indicated by the inter-rater reliability rate (Stemler 2004) of the teachers, both individually and collectively.

The present study has made a contribution in terms of theory, assessment practice

and professional learning. In terms of theory, the present study draws on SFL theory of language (e.g., Halliday 1985a; 1994, Martin 1997, Martin, Christie and Rothery 1987, Rose and Martin 2012). Therefore, the present study has made a contribution to research in terms of taking the particular theory that was developed in an English-dominant society and education environment (ESL context) and experimenting with it in the Indonesian-dominant education (EFL context) environment. The concepts of genre and register and their impacts on language choices helped identify features of successful essays, and informed the design and application of the new marking rubric in the assessment of students' work. The system of Appraisal has been used to systematically explore teachers' accounts in terms of their attitudes and opinions towards the two rubrics used, their marking consistencies and perceived difficulties, and the intervention. Thus, the study has also contributed to a wider application of the Appraisal taxonomy to analysis of interviews and assessment, which was somewhat limited (Thomson, Cleirigh et al. 2006, Tran 2011).

The present study has also contributed to assessment policy and practice. In terms of policy in an institution, the study has looked at assessment that is informed by successful criteria and linguistic principles. In assessment practice, the study is making a contribution in terms of making visible what counts in quality essays.

In terms of the development of professional learning, the study engaged the teachers in a research-based process by which they were actually improving their assessment practice by being involved in this rubric design. In other words, the ideas were not imposed on the teachers from outside, but instead, the teachers have been part of the process and have helped develop and refine the marking rubric.

### **7.2.3 Limitations of the study**

There are two ostensible limitations of the study, which are all related to the methodological issues. The first limitation is the small sample size, which was made up of six EFL teachers and 20 EFL students' essays. Thus, the results cannot be generalised to other cases. However, it has not been the intention of the current case study to make generalisations based on the findings of a small sample. Rather, through the description of the findings, the study has attempted to shed light on the potential of application of SFL-informed rubric on EFL writing, particularly on EAP programs in Indonesia. In a field dominated by Western rhetoric theory, this study highlights the importance of reliability among Indonesian English teachers in marking student essays based on the academically accepted writing in Western world. The second and last limitation is the relatively short time frame for the intervention due to the teachers' work commitments. As a result, the intervention comprised only a four-session (with a total of eight hours) teacher-training program. This has meant that the full scope of the SFL systems could not be meticulously explained to the teachers. Teachers' statements gained from the interview data have indicated the needs to allocating more time to the teacher-training program.

### **7.2.4 Recommendations for further research**

The study has shown that a linguistically informed assessment rubric introduced to teachers through an intervention was able to improve reliability of marking among EAP teachers. However, further research needs to be conducted so as to enable generalisations of the findings. In terms of scale, the replications of the study can be conducted with increased numbers of participants, and applications of assessment

practices to different genres of writing. In terms of timing, it will be useful to carry out further research with a more extensive program. This would establish the validity of SFL-informed criteria in the marking of academic essays in the Indonesian education context.

The relationship between teachers' assessment practices and perceptions is an under-researched area. In order to draw more reliable conclusions about the relationship between teachers' assessment practices and perceptions, there needs to be more research using different methodologies to investigate this issue. A fruitful line for future research might well involve adopting EFL teachers' perceptions and marking practices to broader and higher educational levels such as regional high school and university or even at the state level. Longitudinal studies examining teachers' marking and their perceived reliability are likely to illuminate other hidden aspects of the topic. These longitudinal studies will provide important insights into teachers' capabilities in identifying text quality using functional language analysis and whether or not these identifications can be found in their marking of student essays. Such studies will also provide insights into the teachers' abilities to generate systematic linguistic evidence that supports their judgements of the essays, as argued by Fang and Wang (2011).

### **7.2.5 Concluding remarks**

This thesis addresses the research questions that drive this study. The study has identified and addressed problems in teachers' assessment of student essays by incorporating teachers' opinions into the design of a new assessment rubric based on the coherent model of language informed by SFL. The results presented in the thesis

have shown that the SFL-informed rubric succeeded in improving the reliability of teachers' marking practices. In terms of perceptions, the results of the study have brought about an in-depth understanding of teachers' attitudes and opinions of their assessment practices in terms of capabilities when using the two different rubrics and shows the correlations between teachers' perceptions and their assessment practices.

## References

- Abasi, R. A. (2012). The pedagogical value of intercultural rhetoric: A report from a Persian-as-a-foreign-language classroom. *Journal of Second Language Writing* 21 (2): 195-220.
- Abercombie, N., S. Hill and B. S. Turner (2006). *The Penguin dictionary of sociology*. London: Penguin.
- Arkoudis, S. and K. O'Loughlin (2004). Tensions between validity and outcomes: teacher assessment of written work of recently arrived immigrant ESL students. *Language Testing* 21 (3): 284-304.
- Arksey, H. and P. Knight (1999). *Interviewing for social scientists: An introductory resource with examples*. London: Sage Publications.
- Babbie, E. (2004). *The practice of social research*. (10th ed). Belmont: Thomson.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., B. K. Lynch and M. Mason (1995). Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing* 12 (2): 238-257.
- Bachman, L. F. and A. S. Palmer (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing* 12 (2): 86-107.
- Barkaoui, K. (2010a). Do ESL essay raters' evaluation criteria change with experience? A mixed-methods, cross-sectional study. *TESOL QUARTERLY* 44 (1): 31-57.
- Barkaoui, K. (2010b). Explaining ESL essay holistic scores: A multilevel modeling approach. *Language Testing* 27 (4): 515-535.
- Barnwell, D. (1989). 'Naive' native speakers and judgments of oral proficiency in Spanish. *Language Testing* 6 (2): 152-163.

- Bauer, B. A. (1981). *A study of the reliabilities and cost-efficiencies of three methods of assessment for writing ability.*, (ERIC document reproduction service No. ED 216357).
- Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science.*Madison: The University of Wisconsin Press.
- Belcher, D. (Ed.). (2009). *English for specific purposes in theory and practice.* Ann Arbor: University of Michigan Press.
- Bereiter, C. and M. Scardamalia (1987). *The psychology of written composition.*Hillsdale: Lawrence Elbraum Associates.
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings.*London: Longman.
- Bloor, T. and M. Bloor (2004). *The functional analysis of English (2nd ed).*London: Arnold.
- Boettger, R. K. (2010). Rubric use in technical communication: Exploring the process of creating valid and reliable assessment tools. *IEEE Transactions on Professional Communication* 53 (1): 4-16.
- Bogdan, R. C. and S. K. Biklen (2003). *Qualitative research for education: An introduction to theory and methods.*(4). Boston: Pearson.
- Brannan, D. and T. Bleistien (2002). Novice ESOL teachers' perceptions of social support networks. *TESOL QUARTERLY* 46 (3): 519-541.
- Brossell, G. (1986). Current research and unanswered questions in writing assessment. In K. L. Greenberg, H. S. Weiner and R. A. Donovan (Eds), *Writing assessment.* New York: Longman. 168-182.
- Brown, H. D. (2004). *Language Assessment Principles and Classroom Practices.*New York: Pearson education.
- Brown, J. D. (1991). Do English and ESL faculties rate writing samples differently? . *TESOL QUARTERLY* 25 (4): 587-603.
- Bruce, L. (1996). *Part time English: Introduction to academic writing.* Sydney: Insearch Language Centre.
- Butt, D., R. Fahey, S. Feez, S. Spinks and C. Yallop (2000). *Using functional grammar: An explorer's guide.*(2nd edn). Sydney: National Centre for English Language Teaching and Research.

Cast, B. M. D. (1939). The efficiency of different methods of marking English composition. *British Journal of Education and Psychology*. 9 (3): 257-269.

Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the teaching of English* 18 (1): 65-81.

Christie, F. (2002). The development of abstraction in adolescence in subject English. In M. J. Scleppegrell and M. C. Colombi (Eds), *Developing advanced literacy in first and second languages: Meaning with power*. Mahwah:: Lawrence Elbaum Associates, Inc. 45-66.

Christie, F. and B. Derewianka (2008). *School discourse*. London: Continuum International Publishing Group.

Christie, F. and J. Rothery (1989). Exploring the written mode and the range of factual genres. In F. Christie (Ed.), *Writing in schools*. Geelong: Deakin University Press. 137-169.

Coffin, C. (1997). Constructing and giving value to the past: An investigation into secondary school history. . In F. Christie and J. R. Martin (Eds), *Genre and Institutions: Social processes in the work place and school*. London: Continuum.

Coffin, C. (2000). *history as discourse: Construal of time, cause and Appraisal*. Unpublished doctoral dissertation

Coffin, C. (2004). Arguing about how the world is or how the world should be: the role of argument in IELTS tests. *Journal of English for Academic Purposes* 3 (3): 229-246.

Coffin, C. and A. Hewings (2004). IELTS as preparation for tertiary writing: distinctive interpersonal and textual strategies. In L. J. Ravelli and R. A. Ellis (Eds), *Analysing academic writing: Contextualized frameworks*. London: Continuum.

Coffman, W. E. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement* 3 (2): 151-156.

Cohen, A. and M. Cavalcanti (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom (3rd ed.)*. New York: Cambridge University Press.

Cohen, L., L. Manion and K. Morrison (2000). *Research methods in education*. (6th ed.). London: Routledge.

Connor, U. (2004). Intercultural rhetoric research: Beyond texts. *Journal of English for Academic Purposes* 3 (4): 291-304.

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*.(2). New Jersey: Pearson Education.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches*.(2nd ed.). Thousand Oaks, California: Sage.

Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks: California: Sage.

Crocker, L. (2002). Stakeholders in comprehensive validation of standard-based assessments: A commentary. *Educational Measurement: Issues and Practice* 22 (1): 5-6.

Cronbach, L. J. (1990). *Essentials of psychological testing*.(5th edn). New York: HarperCollins.

Cumming, A. (1989). writing expertise and second language proficiency. *Language Learning* 39 (1): 81-141.

Cumming, A. (1990). Expertise in evaluating second language composition. *Language Testing* 7: 31-51.

Dang, T. K. A. (2002). *Applying a genre-based approach in teaching the writing of the English discussion genre to Vietnamese EFL students at tertiary level*. MEd Thesis, University of Melbourne.Melbourne.

Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley and T. McNamara (1999). *Dictionary of language testing*.Cambridge: Cambridge University Press/ UCLES.

Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL assessment practices in Australia and Hong Kong. . *Language Testing* 21 (3): 305-334.

de Oliveira, L. C. and S. W. Lan (2014). Writing science in an upper elementary classroom: A genre-based approach to teaching English language learners. *Journal of Second Language Writing* 25: 23-39.

Derewianka, B. (2003). Trends and issues in genre-based approaches. *RELC* 34 (2): 133-154.

Derewianka, B. (2007). Using Appraisal theory to tract Interpersonal development in adolescent academic writing. In A. McCabe, M. O'Donnell and R. Whittaker (Eds), *Advances in language and education*. London: Continuum. 142-165.

Derewianka, B. (2012). Knowledge about language in the Australian curriculum: English. *Australian Journal of Language and Literacy* 35 (2): 127-146.

Derewianka, B. and P. Jones (2012). *Teaching language in context*. Melbourne: Oxford University Press.

Dirlam, D. K. (1980). Classifiers and cognitive development. In S. Modgil and C. Modgil (Eds), *Toward a theory of psychological development*. Windsor: NFER Publishing. 465-498.

Dirlam, D. K. and M. Byrne (1978). *Standardized developmental ratings.*, (ERIC document reproduction service No. ED 174629).

Duranti, A. (1997). *linguistic anthropology*. Cambridge: Cambridge University Press.

Duranti, A. (2006). Agency in language. In A. Duranti (Ed.), *A companion to linguistic anthropology*. Oxford: Blackwell Publishing. 451-473.

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing* 14 (2): 88-115.

Eckes, T. (2008). Rater types in writing performance assessment: A classification approach to rater variability. *Language Testing* 25 (2): 155-185.

Eckes, T. (2012). Operational rater types in writing assessment: Linking rater cognition to rater behavior. *Language Assessment Quarterly* 9 (3): 270-292.

Edgeworth, F. Y. (1890). The element of chance in competitive examinations. *Journal of the Royal Statistical Society*. 53 (3): 460-475.

Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. London: Printer Publishers, Ltd.

Eggins, S. (2004). *An Introduction to Systemic Functional Linguistics*. (2nd edn). New York: Continuum.

Elder, C., G. Barkhuizen, U. Knoch and J. von Randow (2007). Evaluating rater responses to an online training program for L2 writing assessment. *Language Testing* 24 (1).

Emilia, E. (2005). *A critical genre-based approach to teaching academic writing in a tertiary EFL context*. Doctor of Philosophy thesis, University of Melbourne - Faculty of Education, University of Melbourne

Emirbayers, M. and A. Mische (1998). What is agency? *American Journal of Sociology* 103: 962-1023.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing* 4 (2): 139-155.

Faez, F. and A. Valeo (2012). TESOL teacher education: Novice teachers' perceptions of their preparedness and efficacy in the classroom. . *TESOL QUARTERLY* 46 (3): 450-471.

Fang, Z. and Z. Wang (2011). Beyond rubrics: Using functional language analysis to evaluate student writing. *Australian Journal of Language and Literacy* 34 (2): 147-165.

Feez, S. (2002). Heritage and innovation in second language education. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives*. Mahwah, New Jersey: Lawrence Elbraum Associate Publishers.

Feldman, M. S. and B. T. Pentland (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly* 48 (1): 94-118.

Feldt, L. S. and R. L. Brennan (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement (3rd ed)*. Washington, DC: The American Council on Education and the National Council on Measurement in Education.

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing* 19: 6-23.

Ferris, D. R., J. Brown, H. Liu and M. Eugenia (2011). Responding to L2 students in college writing classes: Teacher perspectives. *TESOL QUARTERLY* 45 (2): 207-234.

Finson, K. D. and C. K. Ormsbee (1998). Rubrics and their use in inclusive science. *Intervention in School and Clinic* 34 (2): 79-88.

Gall, M. D., J. P. Gall and W. R. Borg (2007). *Educational research: An introduction*. (8th edn). Boston: Pearson.

Gay, L. R. (1985). *Educational evaluation and measurement: Competencies for analysis and application (2nd ed)*. Columbus: C.E Merrill Publishing Company.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.

Gipps, C. (2012). *Beyond testing: Towards a theory of educational assessment*. London: Routledge.

Grabe, W. and R. B. Kaplan (1996). *Theory and practice of writing: An applied linguistic perspective*. New York: Longman.

Gui, M. (2012). Exploring differences between Chinese and American EFL teachers' evaluations of speech performance. *Language Assessment Quarterly* 9 (2): 186-203.

Hafner, J. C. and P. M. Hafner (2003). Quantitative analysis of rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education* 25 (12): 1509-1528.

Halliday, M. A. K. (1975). *Learning how to mean: Explorations in the development of language*. London: Edward Arnold.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Halliday, M. A. K. (1985a). *An introduction to functional grammar*. London: Edward Arnold.

Halliday, M. A. K. (1985b). *Spoken and written language*. Geelong: Deakin University Press.

Halliday, M. A. K. (1994). *An introduction to functional grammar*. (2nd ed). London: Edward Arnold.

Halliday, M. A. K. (1998). Things and relations: Regrammaticising experience as technical knowledge. In J. R. Martin and R. Veel (Eds), *Reading science: critical functional perspectives on discourses of science*. London: Routledge.

Halliday, M. A. K. (2001). Literacy and linguistics: Relationships between spoken and written language. In A. Burns and C. Coffin (Eds), *Analysing English in a global context: A reader*. London: Routledge. 181-193.

Halliday, M. A. K. (2004) The language of science. In J. J. Webster (Ed.) (Eds), *Collected works of M.A.K. Halliday* London: Continuum.

Halliday, M. A. K. and R. Hasan (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Victoria: Deakin University.

Halliday, M. A. K. and C. M. I. M. Matthiessen (1999). *Construing experience through meaning: A language-based approach to cognition* London: Continuum.

Halliday, M. A. K. and C. M. I. M. Matthiessen (2004). *An introduction to functional grammar* (3rd edn). London: Arnold.

Hamp-Lyons, L. (1991a). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation. 241-276.

Hamp-Lyons, L. (1991b). Reconstructing 'academic writing proficiency'. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. New Jersey: Ablex Publishing Corporation. 127-153.

Hamp-Lyons, L. (2001). English for academic purposes. In R. Carter and D. Nunan (Eds), *The Cambridge guide to Teaching English to Speakers of Other Languages*. Cambridge: Cambridge University Press. 126-130.

Hamp-Lyons, L. (2007). Worrying about rating. *Assessing Writing* 12 (1): 1-9.

Hamp-Lyons, L. and A. Davies (2008). The Englishes of English tests: bias revisited. *World Englishes* 27 (1): 26-39.

Hamp-Lyons, L. and B. W. Zhang (2001). World Englishes: issue in and from academic writing assesement. . In H. Flowerdew and M. Peacock (Eds), *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press. 101-116.

Hennink, M., I. Hutter and A. Bailey (2011). *Qualitative research methods*. London: Sage.

Herr, K. and G. L. Anderson (2005). *The action research dissertation: A guide for students and faculty*. London: Thousand Oaks.

Hinds, J. (1987). Reader versus writer responsibility: A new typology. In U. Connor and R. Kaplan (Eds), *Writing across languages: Analysis of L2 texts*. Reading: Addison-Wesley. 141-152.

Hood, S. (2004). Managing attitude in undergraduate academic writing: a focus on the introductions to research reports. In L. J. Ravelli and R. A. Ellis (Eds), *Analysing academic writing: Contextualized frameworks*. London: Continuum.

Hood, S. (2010). *Appraising research: Evaluation in academic writing*. Basingstoke: Palgrave Macmillan.

Hughes, A. (2003). *Testing for language teachers*. (2nd ed.). Cambridge: Cambridge University Press.

- Huot, B. (1990). Reliability, validity, and holistic scoring: what we know and what we need to know. *College Composition and Communication* 41 (2): 201-213.
- Huot, B. (1996). Toward the new theory of writing assessment. *College Composition and Communication* 47 (4): 549-566.
- Hyland, K. (2006). *English for academic purpose: An advanced resource book*. London: Routledge.
- Hyland, K. (2009). *Academic discourse: English in a global context*. London: Continuum International Publishing.
- Hyland, K. (2013). Faculty feedback: Perceptions and practices in L2 disciplinary writing. *Journal of Second Language Writing* 22 (3): 240-253.
- Hyland, K. and E. Anan (2006). Teachers' perceptions of error: The effect of first language and experience. *SYSTEM* 34 (4): 509-519.
- Jacobs, H. L., S. A. Zingraf, D. R. Wormuth, V. F. Hartfield and J. B. Hughey (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.
- Jakeman, V. and C. McDowell (2003). *Insight into IELTS extra: the Cambridge IELTS course*. Cambridge: Cambridge University Press.
- Johns, A. M. (1997). *Text, role and context: Developing academic literacies*. Cambridge: Cambridge University Press.
- Johnson, J. S. and G. S. Lim (2009). The influence of rater language background on writing performance assessment. *Language Testing* 26 (4): 485-505.
- Jonsson, A. and G. Svingby (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review* 2 (2): 130-144.
- Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning* 16 (1): 1-20.
- Keeves, J. P. (Ed.). (1988). *Educational research, methodology, and measurement: An international handbook*. Oxford: Pergamon Press.
- Kervin, L., W. Vialle, J. Herrington and T. Okely (2006). *Research for educators*. South Melbourne: CENGAGE Learning.

- Knapp, P. and M. Watkins (1994). *The genre of arguing* Broadway, NSW: Text Production.
- Knoch, U. (2007). 'Little coherence, considerable strain for reader': A comparison between two rating scales for the assessment of coherence. *Assessing Writing* 12 (2): 108-128.
- Knoch, U. (2008). The Assessment of Academic Style in EAP writing: The Case of the Rating Scale. *Melbourne Papers in Language Testing* 13 (1): 34-67.
- Knoch, U. (2009). Diagnostic Assessment of Writing: A comparison of two rating scales *Language Testing* 26 (2): 275-304.
- Kondo-Brown, K. (2002). A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing* 19 (1): 3-31.
- Kongpetch, S. (2003). *The implication of the genre-based approach on the teaching English writing at the department of foreign languages, Khon Kaen University, in north-eastern Thailand*. Unpublished PhD, University of Technology Sydney
- Kumar, R. (1996). *Research methodology: A step-by-step guide for beginners*. South Melbourne: Longman.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage Publications.
- Lapan, S. D., M. T. Quartaroli and F. J. Riemer (Eds). (2012). *Qualitative reasearch: An introduction to methods and designs*. Hoboken: Jossey-Bass.
- Lee, H. K. (2008). The relationship between writers' perceptions and their performance on a field-specific writing test. *Assessing Writing* 13: 91-110.
- Lee, H. K. (2009a). Native and nonnative rater behavior in grading Korean students' English essays. *Asia Pacific Education Review* 10 (3): 387-397.
- Lee, I. (2008a). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing* 17 (2): 69-85.
- Lee, I. (2008b). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing* 17 (3): 144-164.
- Lee, I. (2009b). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal* 63 (1): 13-22.

- Lee, I. (2010). Writing teacher education and teacher learning: Testimonies of four EFL teachers. *Journal of Second Language Writing* 19 (3): 143-157.
- Lee, S. H. (2006). *The use of interpersonal resources in argumentative/persuasive essays by East-Asian ESL and Australian tertiary students*. Unpublished PhD Thesis, University of Sydney.Sydney.
- Lee, S. H. (2007). An application of multiple coding for the analysis of ATTITUDE in an academic argument. *Linguistics and the Human Sciences* 3 (2): 165-190.
- Lee, S. H. (2010). Attribution in high-and low-graded persuasive essays by tertiary students. *Journal of Functions of Language* 17 (2): 181-206.
- Lee, S. H. (2013). Argument structure as an interactive resource by undergraduate students. *Linguistics and the Human Sciences* 9 (2): 273-299.
- Lemke, J. L. (1992). Interpersonal meaning in discourse: value orientations. In A. Davies and L. J. Ravelli (Eds), *Advances in systemic linguistics: Recent theory and practice*. London: Pinter Publishers. 82-104.
- Lemke, J. L. (1995). *Textual politics: Discourse and social dynamics*.Bristol: Taylor & Francis.
- Lim, G. S. (2009). *Prompt and rater effects in second language writing performance assessment* Unpublished disertation, The Unversity of Michigan
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: what do they really mean to raters? *Language Testing* 19 (3): 246-276.
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective*.Frankfurt: PETER LANG.
- Lynch, B. K. and T. F. McNamara (1998). Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing* 15: 158-180.
- Macken-Horarik, M. (2002). Something to shoot for: A systemic functional approach to teaching genre in secondary school science. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives*. Mahwah: Lawrence Elbraum Associates, Inc. 17-42.
- Martin, J. R. (1984). Language, register and genre. In F. Christie (Ed.), *Children writing: Reader*. Geelong: Deakin University. 21-30.

Martin, J. R. (1985). *Factual writing: Exploring and challenging social reality*. Burwood: Deakin University Press.

Martin, J. R. (1991). Nominalisation in science and humanities: distilling knowledge and scaffolding text. In E. Ventola (Ed.), *Functional and systemic linguistics: Approaches and uses*. Berlin: Mouton de Gruyter. 307-338.

Martin, J. R. (1992). *English text: System and structure*. Philadelphia: John Benjamins.

Martin, J. R. (1993). A contextual theory of language. In B. Cope and M. Kalantzis (Eds), *The powers of literacy: A genre approach to teaching writing*. London: The Falmer Press. 116-136.

Martin, J. R. (1997). Analyzing Genre: Functional parameter. In F. Christie and J. R. Martin (Eds), *Genre and institutions: Social processes in the workplace and school*. London: Cassell.

Martin, J. R. (2000). Beyond exchange: Appraisal system in English. In S. Hunston and G. Thompson (Eds), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. 142-175.

Martin, J. R. (2001). Technicality and abstraction: Language for the creation of specialized texts. In A. Burns and C. Coffin (Eds), *Analysing English in a global context: A reader*. London: Routledge. 211-228.

Martin, J. R. (2005). Analyzing Genre: Functional parameter. . In F. Christie and J. R. Martin (Eds), *Genre and institutions: Social processes in the workplace and school*. London: Continuum International Publishing.

Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education* 20 (1): 10 - 21.

Martin, J. R., F. Christie and J. Rothery (1987). Social processes in education: A reply to Sawyer and Watson. *metaphor* (4): 51-52.

Martin, J. R., C. Matthiessen and C. Painter (1997). *Working with functional grammar*. London: Arnold.

Martin, J. R. and D. Rose (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.

Martin, J. R. and D. Rose (2007). *Working with discourse: Meaning beyond the clause*. London: Continuum.

- Martin, J. R. and D. Rose (2008). *Genre relations: Mapping culture* London: Equinox.
- Martin, J. R. and J. Rothery (1980) *Writing project report: Working papers in linguistics, No. 1.*
- Martin, J. R. and J. Rothery (1981) *Writing project report: Working papers in linguistics, No. 2.*
- Martin, J. R. and J. Rothery (1993). Grammar: Making meaning in writing. In B. Cope and M. Kalantzis (Eds), *The powers of literacy: A genre approach to teaching writing*. London: The Falmer Press. 137-153.
- Martin, J. R. and P. R. R. White (2005). *The Language of Evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- Martin, J. R. and P. R. R. White (2007). *The Language of Evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- Matsuda, P. K., T. Saenkhum and S. Accardi (2013). Writing teachers' perceptions of the presence and needs of second language writers: An institutional case study. *Journal of Second Language Writing* 22 (1): 68-86.
- Matthiessen, C. M. I. M. (1992). Interpreting textual metafunction. In A. Davies and L. J. Ravelli (Eds), *Advances in systemic linguistics: Recent theory and practice*. London: Pinter Publishers. 38-81.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco: Jossey-Bass.
- Miles, D. H. (2003). *The 30-second encyclopedia of learning and performance: A trainer's guide to theory, terminology and practice*. New York: Amacom.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech* 70 (2): 151-167.
- Mohan, B. A. and W. A. Lo (1985). Academic writing and Chinese students: Transfer and developmental factors. *TESOL QUARTERLY* 19 (3): 515-534.
- Montgomery, J. L. and W. Baker (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing* 16 (2): 82-99.

- Moore, T. and J. Morton (1999) *Authenticity in the IELTS academic module writing test: A comparative study of Task 2 items and university assignments*. IELTS Research Report, Vol 2, Paper 4, 64-106.
- Moskal, B. M. and J. A. Leydens (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation* 7 (10): 71-81.
- Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher* 23 (2): 5-12.
- Murphy, R. (1982). A further report of investigations into the reliability of marking of GCE examinations. *British Journal of Education and Psychology*. 52 (1): 58-63.
- Nakamaru (2010). Lexical issues in writing center tutorials with international and US-educated multilingual writers. *Journal of Second Language Writing* 19 (2): 95-113.
- Nakamura, A. (2009). *Construction of evaluative meanings in the IELTS writing: An intersubjective and intertextual perspective*. Doctor of Philosophy thesis, University of Wollongong - Faculty of Education, University of Wollongong
- Norris, J. M. (2008). *Validity evaluation in language assessment*. Frankfurt: Peter Lang.
- North, B. and G. Schneider (1998). Scaling Descriptors for Language Proficiency Scales. *Language Testing* 15 (2): 217-263.
- Nunan, D. (2008). Exploring genre and register in contemporary English. *English Today* 24 (2): 56-61.
- Patton, M. (2002). *Qualitative research and evaluation methods*. (3). Thousand Oaks, CA: Sage Publications.
- Payaprom, S. (2012). *The impact of genre-based approach on English language teaching in an EFL tertiary context in Thailand*. Doctor of Education Thesis, University of Wollongong
- Pecorari, D. and S. D. Shaw (2012). Types of student intertextuality and faculty attitudes. *Journal of Second Language Writing* 21 (2): 149-164.
- Penny, J., R. L. Johnson and B. Gordon (2000). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing* 7 (2): 143-164.

Pihlanto, P. (1994). the action-oriented approach and case study method in management studies. *Scandinavian Journal of Management*. 10 (4): 369-382.

Radnor, H. (2002). *Researching your professional practice: Doing interpretive research*. Buckingham: Open University Press.

Ravelli, L. J. (1988). Grammatical metaphor: An initial analysis. In E. Steiner and R. Valtman (Eds), *Pragmatics, discourse and text*. London: Pinter. 133-147.

Ravelli, L. J. (2005). Getting started with functional analysis of texts. In L. Unsworth (Ed.), *Researching language in schools and communities: functional Linguistic Perspectives*. London: Continuum International Publishing. 27-64.

Ravelli, L. J. and R. A. Ellis (2004). *Analysing academic writing: Contextualized frameworks*. London: Continuum.

Rezaei, A. R. and M. Lovorn (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing* 15: 18-39.

Richards, K. (2003). *Qualitative inquiry in TESOL*. New York: Palgrave Macmillan.

Rose, D. (2005). *Reading to learn-book 2-Assessing writing*.

Rose, D. and J. R. Martin (2012). *Learning to write, reading to learn: Genre, knowledge and pedagogy in the Sydney School*. Sheffield: Equinox.

Rothery, J. (1996). Making changes: developing an educational linguistics. In R. Hasan and G. Williams (Eds), *Literacy in society*. London: Longman. 86-123.

Ruegg, R., E. Fritz and J. Holland (2011). Rater sensitivity to qualities of lexis in writing. *TESOL QUARTERLY* 45 (1): 63-80.

Ryan, K. (2002). Assessment validation in the context of high-stake assessment. *Educational Measurement: Issues and Practice* 21 (1): 7-15.

Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL QUARTERLY* 22 (1): 69-90.

Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing* 25 (4): 465-493.

Schleppegrell, M. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah: Erlbaum.

Scollon, R. (1997). Contrastive rhetoric, contrastive poetics, or perhaps something else? *TESOL QUARTERLY* 31 (2): 352-363.

Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing* 18 (3): 303-325.

Shohamy, E., C. M. Gordon and R. Kraemer (1992). The effects of raters' background and training on the reliability of direct writing tests. *The Modern Language Journal* 76 (1): 27-33.

Silvestri, L. and J. Oescher (2006). Using rubrics to increase the reliability of assessment in health classes. *International Electronic Journal of Health Education* 9: 25-30.

Song, B. and I. Caruso (1996). Do English and ESL faculty differ in evaluating the essays of native English-speaking and ESL students? *Journal of Second Language Writing* 5 (2): 163-182.

Spack, R. (1988). Initiating ESL students into the academic discourse community: How far should we go? *TESOL QUARTERLY* 22 (1): 29-52.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, California: Sage.

Stake, R. E. (2010). *Qualitative research: studying how things work*. New York: The Guilford Press.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation* 9 (4): 1-13.

Stuhlmann, J., C. Daniel, A. Delinger, R. Kenton and T. Powers (1999). A generalizability study of the effects of training on teachers' abilities to rate children's writing using a rubric. *Reading Psychology* 20 (2): 107-127.

Sutojo (1994). *Developing competence in writing in English as a Foreign Language: A study in the teaching of writing of report to junior high school students in Indonesia using a genre-based approach*. Unpublished MEd Thesis, Northern Territory University. Darwin.

Swain, M. (1993). Second language testing and second language acquisition: Is there a conflict with traditional psychometric? *Language Testing* 10: 193-207.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Tedick, D. J. and M. A. Mathison (1995). Holistic scoring in ESL writing assessment: What does an analysis of rhetorical features reveal? In D. Belcher and G. Braine (Eds), *Academic writing in a second language: Essays on research and pedagogy*. Norwood: Ablex Publishing Corporation. 205-230.

Thompson, G. (1996). *Introducing functional grammar*. London: Oxford University Press.

Thomson, E., C. Cleirigh, L. Head and P. Muir (2006). Gardeners' talk: A linguistic study of relationships between environmental attitudes, beliefs and practices. *Linguistics and the Human Sciences* 2 (3): 425-460.

Tran, V. T. H. (2011). *A linguistic study on social attitudes toward the quality issues of postgraduate in Vietnam*. Doctor of Philosophy University of Wollongong. Wollongong.

Tran, V. T. H. and E. Thomson (2008). The nature of 'reporter voice' in a Vietnamese hard news story. In E. Thomson and P. R. R. White (Eds), *Communicating conflict: Multilingual case studies of news media*. London: Continuum. 51-64.

Turner, C. E. and J. A. Upshur (2002). Rating scales derived from student samples: Effects of the scale maker and the student sample on scale content and student scores. *TESOL QUARTERLY* 36 (1): 49-70.

Unsworth, L. (2000). Investigating subject-specific literacies in school learning. In L. Unsworth (Ed.), *Researching language in schools and communities: Functional linguistic perspectives*. London: Casell. 245-274.

Vacc, N. N. (1989). Writing evaluation: Examining four teachers' holistic and analytic scores. *The Elementary School Journal* 90 (1): 87-95.

Vaughan, C. (1991). Holistic assessment: What goes on in the rater's minds? In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. Norwood: Ablex. 111-125.

Ventola, E. (1996). Packing and unpacking of information. In E. Ventola and A. Mauranen (Eds), *Academic writing: Intercultural and textual issues*. Amsterdam: John Benjamins Publishing Co. 153-194.

Weigle, S. C. (1994). Effects of training on raters of ESL compositions. *Language Testing* 11 (2): 197-223.

Weigle, S. C. (1998). Using FACETS to model rater training. *Language Testing* 15 (2): 263-287.

Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approach. *Assessing Writing* 6 (2): 145-178.

Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.

Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing* 16 (3): 194-209.

Weir, C. J. (1990). *Communicative language testing*. Hedfordshire: Prentice Hall International (UK) Ltd.

Weir, C. J. and S. D. Shaw (2007). *Examining writing: research and practice in assessing second language writing*. New York: Cambridge University Press.

White, P. R. R. (2001). *Appraisal*. The Appraisal website 07/07/2010, <http://www.grammatics.com/appraisal/>.

White, P. R. R. (2002). Appraisal. In J. Verschueren, J. Ostman, J. Blommaert and C. Bulcaen (Eds), *The handbook of pragmatics*. Amsterdam: John Benjamins Publishing Company.

White, P. R. R. (2006). Evaluative semantics and ideological positioning in journalistic discourse - a new framework for analysis. In I. Lassen (Ed.), *Mediating ideology in text and image: Ten critical studies*. Amsterdam: John Benjamins. 37-69.

White, P. R. R. (2012). *Appraisal and the resources of intersubjective stance*. The Appraisal website 09/10/2012, <http://www.grammatics.com/appraisal/>.

White, P. R. R. and E. Thomson (2008). The news story as rhetoric: Linguistic approaches to the analysis of journalistic discourse. In E. Thomson and P. R. R. White (Eds), *Communicating conflict: Multilingual case studies of the news media*. London: Continuum. 1-24.

Winke, P. (2011). evaluating the validity of high-stakes ESL test: Why teacher's perceptions matter. *TESOL QUARTERLY* 45 (4): 628-660.

Winke, P., S. Gass and C. Myford (2013). Rater's L2 background as a potential source of bias in rating oral performance. *Language Testing* 30 (2): 231\_252.

Wiseman, C. S. (2012). Rater effects: Ego engagement in rater decision making. *Assessing Writing* 17 (3): 150-173.

Wu, S. M. (2005). *Investigating evaluative language in undergraduate argumentative essays*. Unpublished PhD dissertation, National University of Singapore

Wu, S. M. (2008). Investigating the effectiveness of arguments in undergraduate essays from an evaluation perspective. *Journal of Prospect* 23 (3): 59-75.

Yin, R. K. (1994). *Case study research: Design and methods*. (2nd edn). California: Sage.

Yin, R. K. (2003). *Case study research: Design and methods*. (3rd edn). California: Sage.

Yin, R. K. (2009). *Case study reseach: Designs and methods*. (4th edn). California: Sage.



## PARTICIPATION INFORMATION SHEET FOR EAP TEACHERS

**TITLE: Rater Reliability in the Assessment of Academic Writing: An SFL Analysis of Raters'**

### **Rating and Perceptions**

### PURPOSE OF THE RESEARCH

This is an invitation to participate in a study conducted by Gantina Sitepu as part of his doctoral study at the University of Wollongong. The purpose of the research is to identify problems EAP teachers may find when marking student essays in an Indonesian context, to examine the effect of marking rubric on teachers' consistency in marking, and to analyse teachers' perceptions regarding their consistency and the marking rubrics given.

### INVESTIGATORS

Gantina Sitepu (Student Investigator)  
Faculty of Education  
+61 2-42215249  
[gs939@uowmail.edu.au](mailto:gs939@uowmail.edu.au)

Dr. Pauline Jones  
Faculty of Education  
+61 2 42213322  
[paulinej@uow.edu.au](mailto:paulinej@uow.edu.au)

Dr. Honglin Chen  
Faculty of Education  
+62 2 42213941  
[honglin@uow.edu.au](mailto:honglin@uow.edu.au)

### METHOD AND DEMANDS ON PARTICIPANTS

It would be greatly appreciated if you could consider being a participant in this study for this will mean that you are making a valuable contribution to gaining a better understanding of the issues facing when marking student essays, and how marking rubric could be developed to enhance consistency in marking. If you choose to be included, you will be asked to participate in a three-session of two-hour training on Saturdays. Before the training sessions, you will be asked to complete a Self-Reflection Sheet, which takes no more than 5 minutes, and you will be interviewed for approximately 15-30 minutes at the time most convenience to you. The interviews will be audiotaped and transcribed. The transcript will be send to you so that you can check the accuracy of the account.

### POSSIBLE RISKS, INCONVENIENCES AND DISCOMFORTS

Apart from the 15-30 minutes of your time for the interview and the three-session-training given, I can foresee no risks for you. Your involvement in the study is voluntary and you may withdraw your participation from the study at any time and withdraw any data that you have provided to that point. Refusal to participate in

the study will not affect your relationship with the researcher, the Australia Centre Medan and the University of Wollongong.

#### BENEFITS OF THE RESEARCH

This research is not funded by any organisation or faculty and will be conducted as an independent research study. This research may provide a better understanding of factors affecting consistency in marking academic essays, particularly in Indonesian educational context. Findings from the study will be used for academic publications, such as, thesis, and may also be used in educational journals, book chapters and conference presentations. Confidentiality is assured, and the school and you will not be identified in any part of the research.

#### ETHICS REVIEW AND COMPLAINTS

This study has been reviewed by the Human Research Ethics Committee (Social Science, Humanities and Behavioural Science) of the University of Wollongong. If you have any concerns or complaints regarding the way this research has been conducted, you can contact the UoW Ethics Officer on +61 2 4221 4457. You can also ask me in person or email me if you have any concern about any aspect of the study – [gs939@uowmail.edu.au](mailto:gs939@uowmail.edu.au).

Thank you for your interest in this study.

the study will not affect your relationship with the researcher, the Australia Centre Medan and the University of Wollongong.

#### BENEFITS OF THE RESEARCH

This research is not funded by any organisation or faculty and will be conducted as an independent research study. This research may provide a better understanding of factors affecting consistency in marking academic essays, particularly in Indonesian educational context. Findings from the study will be used for academic publications, such as, thesis, and may also be used in educational journals, book chapters and conference presentations. Confidentiality is assured, and the school and you will not be identified in any part of the research.

#### ETHICS REVIEW AND COMPLAINTS

This study has been reviewed by the Human Research Ethics Committee (Social Science, Humanities and Behavioural Science) of the University of Wollongong. If you have any concerns or complaints regarding the way this research has been conducted, you can contact the UoW Ethics Officer on +61 2 4221 4457. You can also ask me in person or email me if you have any concern about any aspect of the study – [gs939@uowmail.edu.au](mailto:gs939@uowmail.edu.au).

Thank you for your interest in this study.

## Appendix B: Consent Form



University of Wollongong

---

### Consent Form for EAP Teachers

#### **Rater Reliability in the Assessment of Academic Writing: An SFL Analysis of Raters' Rating and Perceptions**

I have been given information about *Rater reliability in the assessment of academic writing: An SFL analysis of raters' rating and perceptions* project of Gantina Sitepu, who is conducting this research as part of a Doctoral degree research supervised by Dr. Pauline Jones and Dr. Honglin Chen from the Faculty of Education at the University of Wollongong.

I have been advised that there will be no burden associated with this research, a part from a 15-minute interview and two-hour training of four sessions. I have had an opportunity to ask Gantina any questions I may have about the research and my participation and understand that I am able to contact him at any time during this research.

I understand that my participation in this research is voluntarily, I am free to refuse to participate and I am free to withdraw from the research at any time. My refusal to participate or withdrawal of consent will not affect my relationship with the researcher and the Australia Centre Medan.

If I have any enquiries about the research, I can contact Gantina Sitepu (+61 2 401193347 or [gs939@uowmail.edu.au](mailto:gs939@uowmail.edu.au)) and Dr. Pauline Jones (+61 2 4221 3322 or [paulinej@uow.edu.au](mailto:paulinej@uow.edu.au)) or Dr Honglin Chen (+61 2 4221 3941 or [honglin@uow.edu.au](mailto:honglin@uow.edu.au)), or if I have any concerns or complaints regarding the way the research is or has been conducted, I can contact the Ethics Officer, Human Research Ethics Committee, Office of research, University of Wollongong on +61 2 4221 4457.

By signing below I am indicating my consent to:

- **Participate in one to one interviews (15-30 minutes)**
- **Participate in the four sessions of two-hour rater-training**

I understand that confidentiality of any information I give is assured and that the data collected from my participation will be used for academic publications, such as

thesis, and may also be used in journal publications, book chapters and conference presentations. I consent for it to be used in that manner.

Signed

Date

.....

...../...../.....

Name (please print)

.....

## Appendix C: Interview Guide



University of Wollongong

---

### Interview guide (pre-training interviews)

1. I understand that you have been teaching EAP and marking student essays for more than three years. During these years, can you tell me about your experience with the assessment criteria you use?
2. Now, I would like to ask you about the difficulties you may have had in using the existing marking criteria. Do you find it difficult to use the existing marking criteria?
3. Do you find yourself consistent in marking student essays? Can you explain?
4. There are two teachers teaching and marking in one EAP class, am I right? Now, if compared to your colleague, do you find yourself stricter/more lenient when marking student essays? Can you give an example?
5. Now I want to ask you about the difficulties you might find in marking student essays. Do you find difficulties in marking your students' essays? Or do you find it easy to mark them? Can you elaborate on that?
6. So, what features of student essays do you find most difficult to mark? Why?
7. What features of student essays do you find easiest to mark? Why?
8. I want to ask you about your definition of a good essay. How do you define a good essay?
9. What are two features of student essays that you find most important?
10. Would you agree if I say that a good essay is the one with less grammatical mistakes? Why/why not?
11. Do you always mark your student essays based on the criteria stated on the (existing) marking rubric?
12. And, do you find each criterion in the (existing) marking rubric has the same importance? Can you elaborate on that?
13. Do you agree if I say that these criteria have helped you be more consistent in rating student essays? Why/why not? Can you give an example?
14. If you could add some additional features on these criteria (show the existing marking criteria), what would you add? Why?

## **Interview guide (post-training interviews)**

1. How do you find your level of consistency in marking student essays after the training given? Can you elaborate on that?
2. How do you find the new marking criteria?
3. Do you find any difficulties in marking student essays using this new marking rubric? Can you elaborate on that?
4. Can you tell me the features of the essays you find most difficult to mark when using this new rubric?
5. And, what about the features of the essays you find easiest to mark?
6. Are the criteria used in this rubric clear and easy to understand?
7. Do you find this new rubric help you marking the content of your student essays?
8. What do you see the main differences between the two marking rubrics?
9. And, if you could add additional features to this marking rubric, what would you add? Why?
10. Do you have any suggestions regarding the rater-training program, so that it can be improved in the future?

## Appendix D: colour coding and abbreviation



University of Wollongong

---

Colour codes for Appraisal systems/subsystems (based on Coffin, 2000)

Pink = Affect

Green = Appreciation

Blue = Judgement- social esteem

Purple= Judgement-social sanction

Red = Graduation

Dark yellow = Engagement-contraction

Light blue = Engagement-expansion

Abbreviation for Appraisal subsystems (based on Martin&White, 2005, p. 71)

+ = positive

- = negative

t = token

neg = negation

sat = *satisfaction*

sec = *security*

reac = *reaction*

comp = *composition*

val = *valuation*

norm = *normality*

cap = *capacity*

ten= *tenacity*

ver = *veracity*

prop = *propriety*

mono = *monogloss*

# Appendix E: Training Materials



University of Wollongong

## Session One

### Introduction to Systemic Functional Linguistics

#### Approaches to Teaching Language

##### Traditional Approach

- Language is a list of rules to be acquired by learners.
- It focuses on correct and incorrect.

##### Functional Approach

- It views language as a set of choices which must be made appropriate to the purpose and the context.
- It concerns with how people use real language for real purposes.

#### Three Functions of Language (Metafunction)

##### 1. ideational Function

Language is used to express ideas or experiences: what is going on (events, activities, behaviours, or state of being (sad, angry, or bored), who or what is involved (people, places, things, concept) and the circumstances surrounding these events (where, when, how), and how these are connected in a sentence.

Example:<sup>1</sup>

Keiko,

This is not a good time to visit Washington. It's very hot. It was 102° yesterday. But the city is still crowded with tourists. The museums here are excellent. I went to the Smithsonian and the National Gallery yesterday – fantastic! And there's a new musical at the Kennedy Center I want to see. But the heat! I don't come to Washington in August.

See you soon,  
Bill

<sup>1</sup> Richards, J.C. (1990). *Interchange: English for international communication*. Cambridge: Cambridge University Press.

Example:<sup>2</sup>

Although some people claim that the country offers a better physical environment with friendlier people, others argue that the social amenities and public transport services are superior in the city.

## 2. Interpersonal Function

Language is used to enact social relationship: to interact with other people, to negotiate relationship and to express opinions and attitude.

Example:<sup>3</sup>

| Ways of Interacting                           | Types of Clause          | Examples                                        |
|-----------------------------------------------|--------------------------|-------------------------------------------------|
| To give information                           | Statement (Declarative)  | Sideways has great coffee.                      |
| To ask for information                        | Question (Interrogative) | Who makes the best coffee?                      |
| To ask someone to do something                |                          | Could you make a cup of coffee?                 |
| To encourage someone to think about something |                          | Have you ever wondered about starts in the sky? |
| To make an offer                              |                          | Would you like a cup of coffee?                 |
| To get something done directly                | Command (Imperative)     | Get me a coffee!                                |

Example:<sup>4</sup>

- 'Let me have your mobile – I need to call my husband.'
- 'I wonder if you'd mind awfully lending me your mobile – I need to call my husband.'

---

<sup>2</sup> Jordan, R. R. (1990). *Academic writing course*. Surrey: Thomas Nelson and Sons Ltd.

<sup>3</sup> Droga, L & Humprey, S. (2005). *Grammar and meaning: An introduction for primary teachers*. Berry: Target Text.

<sup>4</sup> Nunan, D. (2008). Exploring genre and register in contemporary English. *English Today*, 24 (2): 56-61.

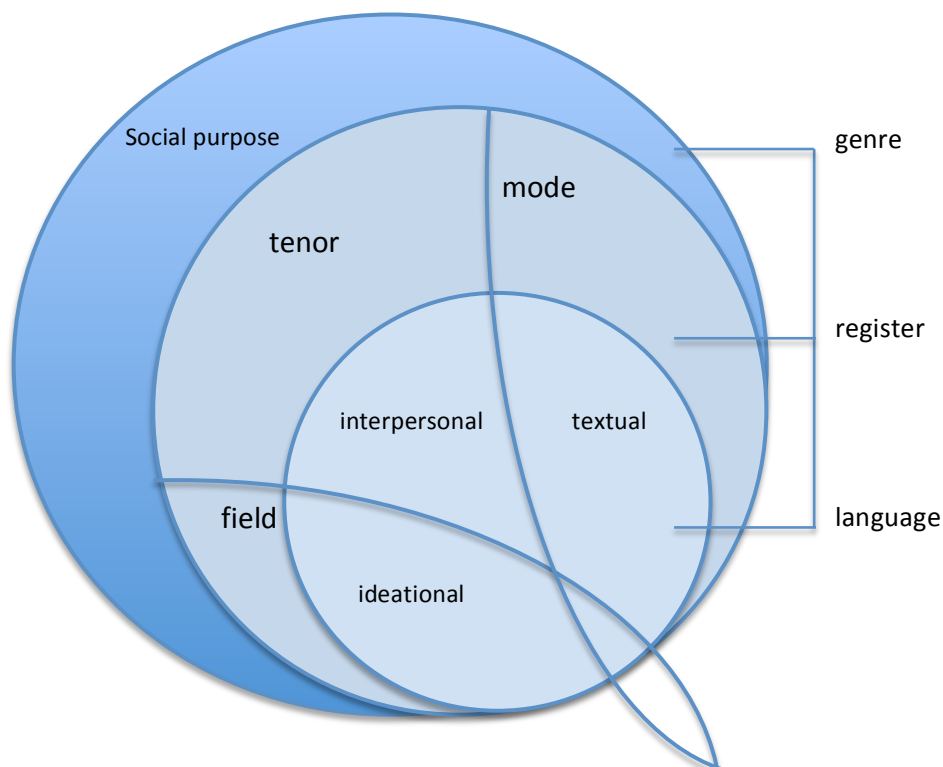
### 3. Textual Function

Language is used to construct logical and coherent texts: how information is organised in a text and how connections are made across a text.

Example: <sup>5</sup>

There are several stages in the manufacturing of **tea**. First of all, the **top buds** and the **two leaves** immediately below are plucked from each plant. In the next stage, the **fresh leaves** are spread on racks or shelves and allowed to wither for between eighteen and twenty-four hours. The next step is to feed the **withered leaves** into a rolling machine for up to three hours in order to release the juices. After this, the **leaves** are fermented on trays before being fired in a furnace for thirty to forty minutes. Then, the **dried leaves** are cut, sifted, sorted and graded. Subsequently, the **tea** is packed in tea chests so that **it** can be exported. After being tasted by experts the **tea** is auctioned. In the final stage, the **tea** is put in packets and sold in shops.

### Language in relation to Register and Genre<sup>6</sup>



<sup>5</sup> Bruce, L. (1996). *Part Time English: Introduction to academic writing*. Sydney: Insearch Language Centre.

<sup>6</sup> Christie, F. & Martin, J. R. (1997). *Genre and institutions: Social processes in the workplace and school*. London: Cassell.

### **Genre and Social Purpose'**

Genre<sup>8</sup> reflects the cultural context within which they are constructed (the context of culture of a text). It is the specific purpose of a text and how the text is organised to achieve the purpose. Some basic educational genres and their sub-genres are explicated below:

| Social purpose                                                   | Genre              | Sub-types                                                                                                          | Example                                                                                                  |
|------------------------------------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| To provide information about a particular person, place or thing | Description        | Objective description                                                                                              | My Family                                                                                                |
|                                                                  |                    | Literary description                                                                                               | The Old Banyan Tree                                                                                      |
| To provide information about a class of things                   | Information report | Descriptive report                                                                                                 | The Rain Forest                                                                                          |
|                                                                  |                    | Taxonomic report<br>- Class/subclass<br><br>- Part/whole<br><br>- Compare/contrast report<br><br>Historical report | Different Types of Planes<br>The Parts of a Clock<br><br>Bats and Birds<br><br>Dinosaurs                 |
| To tell someone how to do something                              | Procedure          | Instructions                                                                                                       | Making Nasi Goreng                                                                                       |
|                                                                  |                    | Experiment<br><br>Directions<br><br>Regulations                                                                    | Changing solids to Liquids<br>How to Get to Hyde Park<br>Our Class Rules                                 |
| To tell what happened                                            | Recount            | Personal recount                                                                                                   | My Holiday in Sydney<br>Thief Steals Car                                                                 |
|                                                                  |                    | Factual recount<br><br>Biographical recount<br><br>Autobiographical recount<br>Historical recount                  | The Life of Gandhi<br>My Childhood<br><br>The Qin Dynasty                                                |
| To explain how or why a phenomenon takes place                   | Explanation        | Sequential explanation                                                                                             | Life Cycle of a Frog                                                                                     |
|                                                                  |                    | Casual explanation<br><br>Factorial explanation<br><br>Consequential explanation                                   | Why Hot Air Rises<br><br>The Causes of Erosion<br>The Effects of Poverty<br>Why Dinosaurs Became Extinct |

<sup>7</sup> Derewianka, B. (2003). *Trends and issues in genre-based approaches*. *RELC*, 34 (2), 133-154.

<sup>8</sup> Martin, J. (2001). Language, register and genre. In A. Burns & C. Coffin, (eds). *Analysing English in a global context*. London: Routledge.

### **Genre and Social Purpose'**

Genre<sup>8</sup> reflects the cultural context within which they are constructed (the context of culture of a text). It is the specific purpose of a text and how the text is organised to achieve the purpose. Some basic educational genres and their sub-genres are explicated below:

| Social purpose                                                   | Genre              | Sub-types                                                                                                          | Example                                                                                                  |
|------------------------------------------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| To provide information about a particular person, place or thing | Description        | Objective description                                                                                              | My Family                                                                                                |
|                                                                  |                    | Literary description                                                                                               | The Old Banyan Tree                                                                                      |
| To provide information about a class of things                   | Information report | Descriptive report                                                                                                 | The Rain Forest                                                                                          |
|                                                                  |                    | Taxonomic report<br>- Class/subclass<br><br>- Part/whole<br><br>- Compare/contrast report<br><br>Historical report | Different Types of Planes<br>The Parts of a Clock<br><br>Bats and Birds<br><br>Dinosaurs                 |
| To tell someone how to do something                              | Procedure          | Instructions                                                                                                       | Making Nasi Goreng                                                                                       |
|                                                                  |                    | Experiment<br><br>Directions<br><br>Regulations                                                                    | Changing solids to Liquids<br>How to Get to Hyde Park<br>Our Class Rules                                 |
| To tell what happened                                            | Recount            | Personal recount                                                                                                   | My Holiday in Sydney<br>Thief Steals Car                                                                 |
|                                                                  |                    | Factual recount<br><br>Biographical recount<br><br>Autobiographical recount<br>Historical recount                  | The Life of Gandhi<br>My Childhood<br><br>The Qin Dynasty                                                |
| To explain how or why a phenomenon takes place                   | Explanation        | Sequential explanation                                                                                             | Life Cycle of a Frog                                                                                     |
|                                                                  |                    | Casual explanation<br><br>Factorial explanation<br><br>Consequential explanation                                   | Why Hot Air Rises<br><br>The Causes of Erosion<br>The Effects of Poverty<br>Why Dinosaurs Became Extinct |

<sup>7</sup> Derewianka, B. (2003). *Trends and issues in genre-based approaches*. RELC, 34 (2), 133-154.

<sup>8</sup> Martin, J. (2001). Language, register and genre. In A. Burns & C. Coffin, (eds). *Analysing English in a global context*. London: Routledge.

|                                                     |                |                                                      |                                                                                  |
|-----------------------------------------------------|----------------|------------------------------------------------------|----------------------------------------------------------------------------------|
|                                                     |                | Exploration                                          |                                                                                  |
| To explore the human condition through storytelling | Story genres   | Narrative<br>Moral tale or fable<br>Anecdote         | The Disappearing dogs<br>The Fox and the Crow<br>When I Went to the Dentist      |
| To respond to a literary text or artistic work      | Response genre | Personal response<br>Review<br>Interpretation        | My Favourite TV show<br>Novel of the Year<br>Crouching Tiger as a Metaphor       |
| To mount an argument                                | Exposition     | Critical response<br>Persuasion<br>Discussion/debate | Is Rap Really Subversive?<br>War is Immoral<br>The Pros and cons of Living Alone |

### Register<sup>9</sup>

Register is the context of situation of a text which examines the relationship between a text and the situational context surrounding the creation of the text. There are three variables (Field, Tenor and Mode) influencing the types of language used in a text

|              |                                                                                                                                   |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------|
| <b>Field</b> | The subject matter of a text, what is going on, what the text is all about                                                        |
| <b>Tenor</b> | The personal relationship between the individuals involved in an activity (the degree of power relationship between interactants) |
| <b>Mode</b>  | The channel of communication being used                                                                                           |

### The Field Continuum

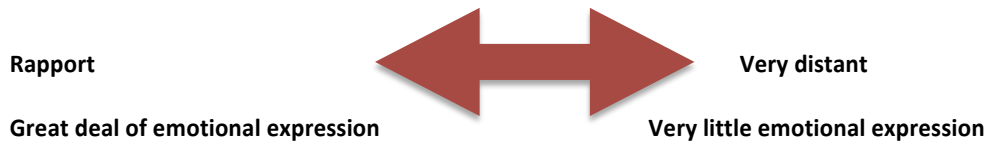


### The Tenor Continuum

Equal status

Great different status

<sup>9</sup>Martin, J. (2001). Language, register and genre. In A. Burns & C. Coffin,(eds). *Analysing English in a global context*. London: Routledge.  
Nunan, D. (2008). Exploring genre and register in contemporary English. *English Today* 94, 24 (2), 56-61.



### The Mode Continuum<sup>10</sup>



### Mixing Register Variable = Different Text Types

- Event: As they sit watching television, a sixteen year-old girl is discussing with her mother what to buy her father for his birthday.
- Here the field is 'buying a gift'. The tenor is mother and daughter. The mode is face-to-face.

- Variation 1: A sixteen year-old girl call her mother to discuss what to buy her father for his birthday.
- Here the field and tenor remain the same, but the mode has changed from face-to-face to telephone conversation.

- Variation 2: a sixteen year-old girl is chatting with her boyfriend about what to buy his father for his birthday.
- Field?
- Tenor?
- Mode?

## Session Two

### Genre and Register Concept activity

|                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                      |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>A: Yes Please<br/> B: Can I have those two?<br/> A: Yes. This one's fifty five and this one's thirty five.<br/> B: And have you got ..... ?<br/> A: Yes. How many would you like?<br/> B: I'll take four<br/> A: Right. That's four dollars ninety altogether.<br/> B: Here you are.<br/> A: Thankyou.<br/> B: Thankyou.</p> | <p>What's the context of the text?<br/> What kind of a text is it? (genre)<br/> What are the stages of the text?<br/> What is it about? (field)<br/> Who is involved? (tenor)<br/> Mode of communication? (mode)</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

<sup>10</sup> Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. London: Pinter.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>The Classification of Birds<sup>11</sup></b></p> <p>Birds are instantly recognisable creatures. Perhaps it is their ability to fly that causes this. Some people might consider that their shape was the most distinguishing feature. Everyone however agrees upon the characteristics that a bird possesses: two wings, feathers, two legs, a toothless bill or beak, warm blood and it lay eggs.</p> <p>The modern system of classifying birds is like a pyramid, with the base formed by 8514 different <i>species</i>. A convenient definition of <i>species</i> is: an interbreeding group of birds which do not normally mate with other such groups.</p> <p>The next division above the species is <i>genus</i>, a group of species showing strong similarities. The scientific name of a bird gives <i>genus</i> first, then the species. Thus, the scientific (latin) name of the golden eagle is 'Aquila chrysaetos' (eagle, golden). When there are strong points of similarity between one genus and another, these related genera are grouped together and are said to belong to the same <i>family</i>. The names of 215 families of birds always end in 'idae'. The golden eagle, for instance, is one of the 'Falconidae' (falcon family).</p> <p>Families with broadly similar characteristics are grouped together into 27 orders, whose names end in 'iformes'. Like the golden eagle falls into the order of 'Falconiformes' (falcon-like birds). The largest order is 'Passeriformes' or perching bird.</p> | <p>What is the purpose of the text?</p> <p>What genre does the text belong to?</p> <p>How is the text organised?</p> <p>What are the distinctive language features of the text?</p> <p>Field?</p> <p>Tenor?</p> <p>Mode?</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### Exercise 2: clause complex and lexical density

Look at the clauses on each paragraph, how are they constructed?

How many clauses are there in each paragraph?

How many sentences consist of single clauses?

How many sentences consist of two or more clauses combination?

Underline all the noun, verbal and adverbial groups in the passage above.

How many noun groups are there in a clause?

What kinds of nouns are there? (pronouns, proper, abstract, nominalisation)

In the following clause "A convenient definition of *species* is: an interbreeding group of birds which do not normally mate with other", how many lexical words can you find? How many noun groups? What is the verb used to connect the noun groups?

<sup>11</sup> Jordan, R. R. (1990). *Academic writing course*. Surrey: Thomas Nelson and Sons Ltd.

### Exercise 3: Flow of information or message

Underline either the first adverbial or noun groups you find on each clause and name it Theme.

And then, underline the rest part of the same clause after the Theme and name it Rheme.

How many of Themes are repeated using synonym or meronym or pronoun?

Can you find clauses in the passage where the Rheme was used as the Theme in the subsequent clause?

Can you identify the flow of information? How does the message delivered to create cohesive and clear flow of information?

#### Is living in the country better than living in the city?<sup>12</sup>

In discussing whether living in the country is preferable to living in the city, it is necessary to look at the advantages of both. Although some people claim that the country offers a better physical environment with friendlier people, others argue that the social amenities and public transport services are superior in the city.

There are several reasons why country-life is superior to city-life. Firstly, the physical environment in the country is cleaner, quieter and more aesthetically pleasing than that in the city. Cities lack space and greenery and are very noisy. The small number of parks and gardens can hardly compensate for the unsightly concrete buildings, noisy crowds and traffic which dominate the city landscape. In addition, the people in the country are usually friendlier than those in the city. While meeting people in cities can be difficult, in the country it is easier to establish a social life in the community. This is because a sense of community spirit exists.

On the other hand, there are a number of advantages to living in the city. One advantage is that cities have better public transport services. Thus, the need for private transport is reduced. Public services are not as well-developed in the country as they are in the city and, consequently, it is necessary to have private a private car. Furthermore, the social amenities in the city are far superior to those in the country. Whereas theatres,

What is the purpose of the text?

What genre does the text belong to?

How is the text organised?

What are the distinctive language features of the text?

Field?

Tenor?

Mode?

<sup>12</sup> Bruce, L. (1996). *Part Time English: Introduction to academic writing*. Sydney: Insearch Language Centre.

|                                                                                                                                                                                                                                    |  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <p>restaurants, sports clubs and other entertainment facilities are usually abundant in cities, they are rather limited in the country. Therefore, it is necessary to travel to the city to take advantage of such facilities.</p> |  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

|                                                                                                                                                                                                                                                                                             |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <p>In conclusion, although there are some obvious advantages to living in the country, the social facilities and public services that exist in cities are far superior to those in the country. Therefore, it can be said that living in the city is better than living in the country.</p> |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|

**Exercise 4: clause complex and lexical density**

Look at the clauses on each paragraph, how are they constructed?

How many clauses are there in each paragraph?

How many sentences consist of single clauses?

How many sentences consist of two or more clauses combination?

Underline all the noun, verbal and adverbial groups in the passage above.

How many noun groups are there in a clause?

What kinds of nouns are there? (pronouns, proper, abstract, nominalisation)

In the following clause “Firstly, the physical environment in the country is cleaner, quieter and more aesthetically pleasing than that in the city.” how many lexical words can you find? How many noun groups? What is the verb used to connect the groups?

**Exercise 5: Flow of information or message**

Underline either the first adverbial or noun groups you find on each clause and put number one (1) on it. And then, put number two (2) on the rest part of the same clause.

How many of number 1s are repeated using synonym or meronym or pronoun?

Can you find in the passage where the rest part of the clause (number 2) was used as the head noun groups in the subsequent clause?

<sup>13</sup>Globalisation is a mixture of positive elements and negative elements. It have a negative impact of globalisation on developing countries.

First, junk food or fast food is a result of globalisation that can make people’s healthy decrease, it meant can make people get diseases. For example, people can get obesity, diabetes, and liver. In fact, in America children have given burger and coke for their breakfast by their parents. This condition means many children get obesity because of this condition.

Second, many factories that build by government or businessmen for their own

<sup>13</sup> The essay is written by a student participant identified as EAPs3

benefits besides they said that for decreasing jobless in developing countries. However, many factories that built and their waste that make global warming and pollution. For example, paper factory that produce paper need trees material to make paper so people have to cut trees and owned their benefits but they do not think about effect of this condistion. It can make flood, global warming, smoke of factory cause air pollution.

Third, era globalisation is era that many technology can make people easier. For example, computers and internet that can make people easier. People can sell things and their needs in Internet. This kind of things can make people more convenience but people can get stress or depression easier because when people cannot have that facilities they can easier get stress or depression.

In conclusion, globalisation have a negative side that have to tackle besides it have positive side on it so it is depend on how you look at it or face it.

<sup>14</sup>Some people argue that the state does *not have the right* to make parents immunise their children. However, I feel the question is not whether they should immunise but whether, as members of society, they *have the right not to*.

Preventative medicine has proved to be the most effective way of reducing the incidence of fatal childhood diseases. As a result of the widespread practice of immunising young children in our society, *many lives have been saved and the diseases have been reduced to almost zero*. In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had immunity, *the diseases spread easily*. Diseases such as dysentery were the result of poor hygiene but *these have long been eradicated since the arrival of good sanitation and clean water*. Nobody would suggest that *we should reverse this good practice now* because dysentery has been wiped out.

*Serious diseases such as polio and smallpox have also been eradicated* through national immunisation programmes. In consequence, children not immunised are far less at risk than they would otherwise be. Parents choosing not to immunise are *relying on the fact that the diseases have already been eradicated*. If the number of parents choosing not to immunise increased, there would be *a similar increase in the risk of the diseases returning*.

Immunisation is *not an issue like seatbelts which affect only the individual*. A decision not to immunise will have *widespread repercussions* for the whole of society and for this reason, I do *not* believe that individuals *have the right* to stand aside in this disease-free society. In my opinion immunisation should be obligatory.

---

<sup>14</sup> Jakeman, V. & McDowell, C. (2003) *Insight into IELTS extra: the Cambridge IELTS course*. Cambridge: Cambridge University Press. 185

**Exercise 6: Compare the two essays above (Globalization and Immunization), in terms of the following:**

Purpose and stages used to achieve it (genre)

Topics or ideas and how they are combined in a sentence (field)

Formality and choice of words in evaluations (tenor)

Organization of messages or flow of ideas (mode)

Which essay has higher quality? Why?

Discuss your findings with your neighbor.

## Session Three

### APPRAISAL

In the Systemic Functional Linguistics (SFL),<sup>15</sup> Appraisal can be defined as a discourse semantic tool of analysis of interpersonal meaning and is concerned more particularly with the language of evaluation, attitude and emotion. Appraisal is regionalised into three interacting domains – *Attitude*, *Engagement* and *Graduation*.

#### **Attitude**

Attitude<sup>16</sup> refers to the resources used to make either positive or negative evaluation of phenomena. Attitude can be divided into three categories. These are: **Affect** (emotion): evaluation by means of the writer/speaker indicating how they are emotionally disposed to the person, thing, happening or state of affairs.

Example:

I love jazz.

This new proposal by the government terrifies me.

#### **Exercise 1**

Underline the resources of Affect used in the text below.

<sup>17</sup>“Well, I’m not worried about Jack’s passion for guns, I just find it boring! He’ll sit there engrossed for hours cleaning them lovingly and trying to impress me with what he knows about them. He’s more proud of his guns than he is of his

<sup>15</sup> Martin, J. R. & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.

<sup>16</sup> Martin, J. R. & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan

<sup>17</sup> Droga, L. & Humprey, S. (2002). *Getting Started With Functional Grammar*. Berry: Target Texts.

kids! I get fed up ... but I trust him with them ... he just enjoys shooting targets and he likes the company of other shooters. I used to be scared of them but now I'm happy he's got friends he's comfortable with."

**Judgement** (ethics): normative assessments of human behaviour typically making reference to rules or conventions of behaviour.

Example:

He corruptly agreed to accept money from those bidding for the contract.

Our new classmate seems rather eccentric.

## Exercise 2

Underline the resources of Judgement in the text below.

<sup>18</sup>The fortunate regulars at The Royal Hotel are today thankful for the heroic efforts of their drinking friend, Kevin Lee, last night. The drama began when an armed gunman entered the bar and began to shout incoherently at the barman. Then, he started shooting madly around the bar. While most of his friends panicked, Kevin kept his cool and cleverly distracted the gunman by throwing a broken glass at the light bulb overhead. Then, singlehandedly, he overpowered the man. Kevin's friends say that, but for his bravery and foresight, they could all be dead.

**Appreciation** (aesthetics): assessments of the form, appearance, composition, impact, significance, etc of human artefacts, natural objects as well as human individuals (but not of human behaviour) by reference to aesthetics and other systems of social value.

Example:

It was a captivating performance.

He has grown a deeply disturbing moustache.

## Exercise 3

Underline the resources of Appreciation in the texts below.

<sup>19</sup>Fox is an interesting picture book written by Margaret Wild and illustrated by Ron Brooks. It is a story of a friendship between a magpie and a dog, which is challenged by an evil fox.

The illustrations in the book are very effective because they seem quick and rough but they are really sophisticated and full of texture. Oil and chalk and charcoal are used effectively to give a bushy feeling. The writing seems like a draft but this matches with the outback pictures. The writing and pictures work together for the exact effect that the writer and illustrator wanted.

Fox is a beautiful and meaningful story, suitable for all ages. Although it seems to be so Australian, the story reminds us of traditional European fables.

---

<sup>18</sup> Droga, L. & Humprey, S. (2002). *Getting Started With Functional Grammar*. Berry: Target Texts.

<sup>19</sup> Droga, L. and Humprey, S. (2003). *Grammar and meaning: An introduction to primary teachers*. Berry: Target Texts.

<sup>20</sup>[*Daily Mail* (London) June 20, 2001: p. 3]

Main Head: **Was Sophie's choice the day 's fashion faux pas?**

Small Head: **Bold and bright, the Countess's rainbow display at Royal Ascot**

THE start of Royal Ascot yesterday ensured a rainbow of fashionable colours on the racecourse. And most, it seemed, beamed out from the Countess of Wessex's striking outfit.

Sophie arrived at the Berkshire course in a striped confection that ensured fashion rather than racing was the main talking point in the enclosures. While the Queen settled for restrained coffee and cream, and the Queen Mother her favourite lilac, Sophie's choice ensured maximum attention. Below, two writers give their verdict on the suitability of the suit designed to stand out in a crowd.

**YES Says D'Argy Smith.**

**NO ONE ever accused Sophie Wessex of being a class fashion act.**

At her sartorial best, she was a pallid imitation of Princess Di. She's never had the body, the legs or the style imagination to look as stunning as Diana. But she usually passed muster well, at least, didn't offend.

Yesterday at Ascot, however, she appeared to have taken leave of her fashion senses. Her suit of large horizontal bands of green, orange, pink and blue interspersed with bands of raffia was topped off with an English country wedding cream high hat, complete with wide brim and flower.

You had to sigh at the awfulness of it all. What on earth was she thinking when she decided 'This suit is for me'? And what sort of person would let her walk out of a store looking like that? Perhaps it's an act of rebellion that she is not prepared to lie low after her recent PR indiscretion.

Who knows? But the colours certainly gave me a headache.

Marcelle D'Argy Smith is a former editor of *Cosmopolitan* magazine

### **Graduation<sup>21</sup>**

One of the distinguishing features of Attitude is that the meanings can be graded up or down. The resources used to grade feelings, judgements and assessments operate as another system within Appraisal, called Graduation. The resources of Graduation can be described in terms of Force and focus.

- 1) Force – the grading of meanings from low to high intensity. Force can be realised explicitly through individual lexical items or implicitly through intensifying or toning down terms which already have a core ideational and interpersonal meaning.

---

<sup>20</sup> White, P. R. R. (2001). Appraisal. [www.grammatics.com/appraisal](http://www.grammatics.com/appraisal). Accessed on July 07, 2010.

<sup>21</sup> Droga, L. & Humprey, S. (2002). *Getting Started With Functional Grammar*. Berry: Target Texts.

- 2) Focus – the grading of meanings as more or less precise by making them sharper or softer.

| FORCE            | KEY REALISATIONS                                                                    | EXAMPLES                                                                                                                                                                                                                        |
|------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Explicit grading | Graders: Adverbials<br>Adjectivals                                                  | He has been slightly odd lately.<br>He is a total nerd.                                                                                                                                                                         |
|                  | Measures of amount,<br>time,<br>distance                                            | There's not much left.<br>He jumped up immediately.<br>He was a long way off.                                                                                                                                                   |
|                  | Repetition                                                                          | It was awful, just awful.                                                                                                                                                                                                       |
|                  | Expletive or swearing                                                               | It was bloody awful!                                                                                                                                                                                                            |
| Implicit grading | Lexical items with<br>grading added to core<br>meanings<br><br>(including metaphor) | <div> <div>Core meaning</div> <div>Graded</div> <div> </div> <div> <div>meaning</div> <div>Like</div> <div>Love</div> <div>Adore</div> </div> <div> <div>He's taken my fancy</div> <div>I'm head over heels</div> </div> </div> |

| FOCUS               | EXAMPLES                                                                                        |
|---------------------|-------------------------------------------------------------------------------------------------|
| Sharpened           | Mahmud is a true friend.<br>Jimmy Hendrix is a true guitar legend.<br>Let's make a clean break. |
| Softened or blurred | It was kind of scary.<br>Bob is a sort of friend.                                               |

#### Exercise 4

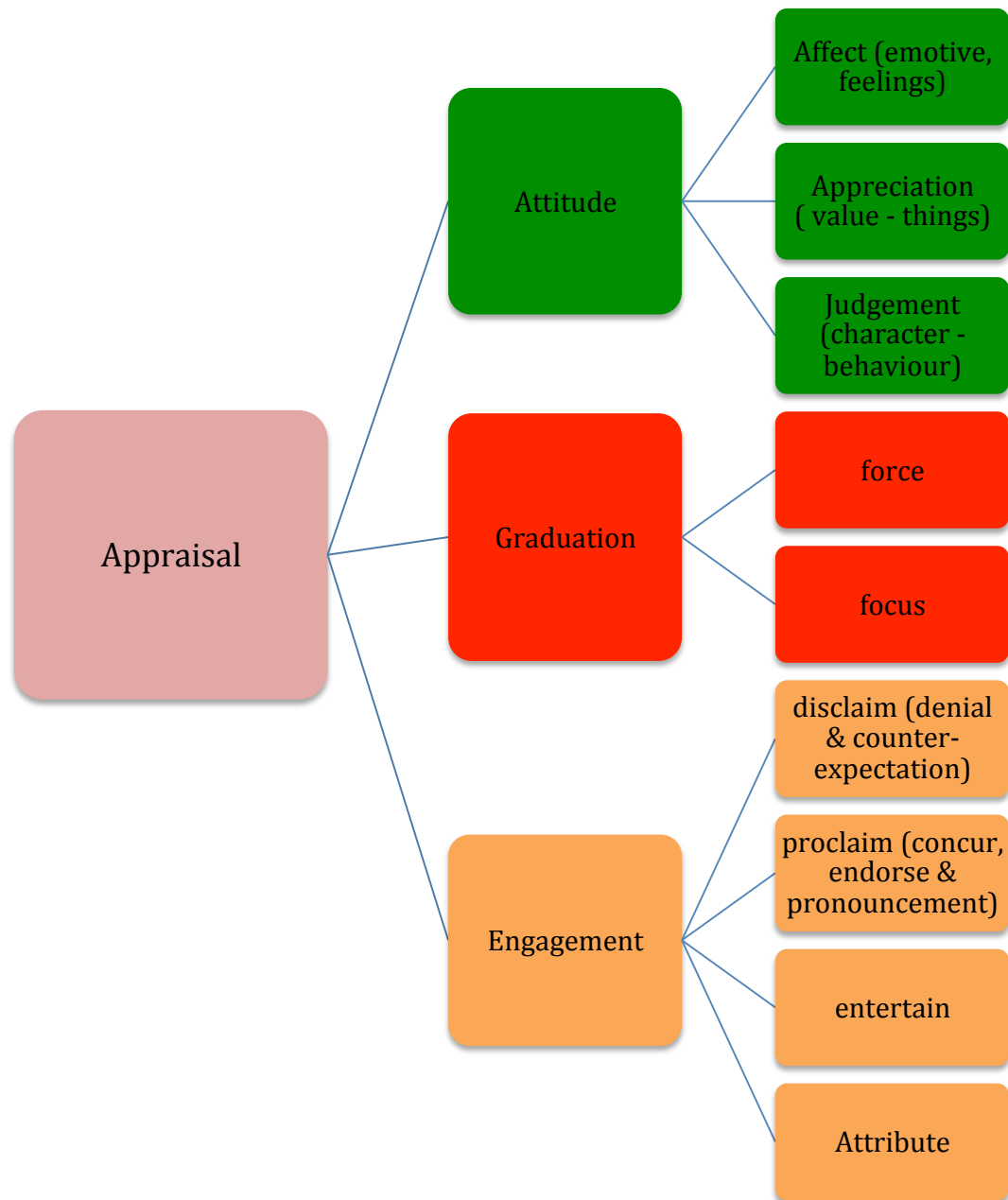
Put a circle on the Graduation elements identified in the four texts above (see exercises 1-3).

#### <sup>22</sup>Engagement

Engagement refers to the diverse range of resources by which speakers/writers adjust and negotiate the arguability of their utterances, propositions or proposals. Key engagement resources include meaning that can be grouped together under the following headings:

<sup>22</sup> White, P. R. R. (2001). Appraisal. [www.grammatics.com/appraisal](http://www.grammatics.com/appraisal). Accessed on July 07, 2010.

## Diagrammatical summary of Appraisal resources: Attitude, Graduation and Engagement



### Exercise 5

Work in pairs and identify all the Appraisal resources in the sentences below

1. <sup>23</sup>Jamie and Adam were disappointed with the movie.
2. The action sequences really gruesome.

<sup>23</sup> Droga, L. & Humprey, S. (2003). *Grammar and meaning: An introduction to primary teachers*.  
Berry: Target Texts.

**1. Disclaim:** includes Denial and Counter-Expectation

- **Disclaim: Denial**

Example:

*The action won't damage the trust between the President and his bodyguards*

- **Disclaim: Counter-Expectation**

Example:

*Amazingly/Bizarrely, this damaged the trust between the President and his bodyguards*

**2. Proclaim:** includes Expectation and Pronouncement

- **Proclaim: Concur**

Example:

*The action will, of course, damage the trust between President and bodyguard.*

*Predictably, the action damaged the trust between President and bodyguard.*

- **Proclaim: Pronouncement**

Example:

*I contend that the action will damage trust.*

*The facts of the matter are that the action damaged the trust between President and body guard.*

*The action undeniably damaged the trust.*

**3. Entertain**

Example:

*It seems that this damaged the trust.*

*This may damage the trust.*

*This probably damaged the trust.*

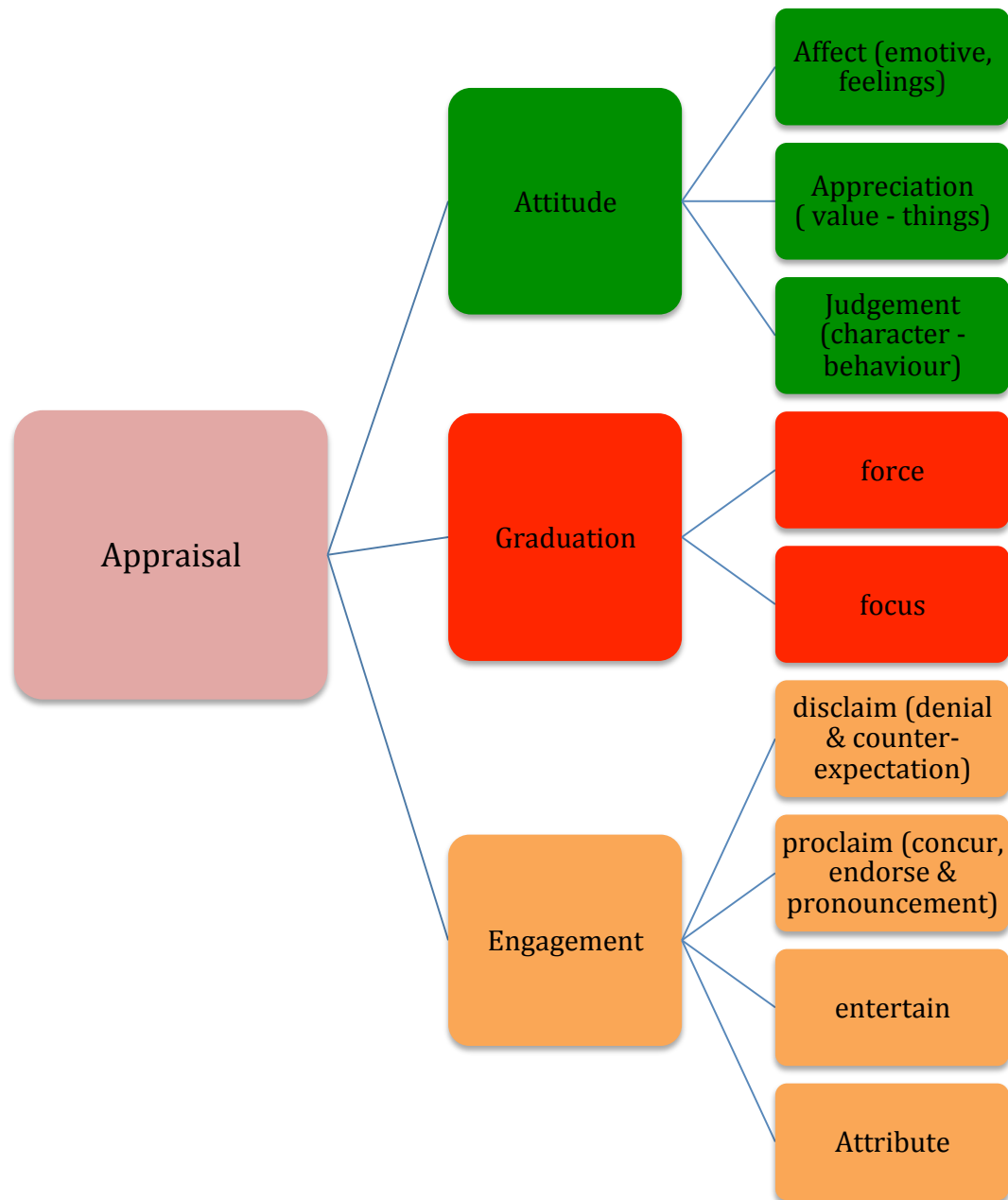
**4. Attribute:** attribution occurs when a writer or speaker (author) refers to the words of thoughts of some outside source.

Example:

*The head of Clinton's security division says this will damage trust.*

*As a number of security experts have indicated, this will damage the trust between President and bodyguard.*

## Diagrammatical summary of Appraisal resources: Attitude, Graduation and Engagement



### Exercise 5

Work in pairs and identify all the Appraisal resources in the sentences below

1. <sup>23</sup>Jamie and Adam were disappointed with the movie.
2. The action sequences really gruesome.

<sup>23</sup> Droga, L. & Humprey, S. (2003). *Grammar and meaning: An introduction to primary teachers*.  
Berry: Target Texts.

3. I heard some academics say that the characters are one-dimensional.
4. I know some people on this forum have said that Harry Potter is too simplistic.
5. I thought the way Snape's true character was revealed was brilliant.

### Exercise 6

Work individually and identify all the Appraisal resources in the text below. After that discuss your findings with your neighbor.

Should parents be obliged to immunise their children against common childhood diseases? Or do individuals have the right to choose not to immunise their children?

<sup>24</sup> Some people argue that the state does *not have the right* to make parents immunise their children. However, I feel the question is not whether they should immunise but whether, as members of society, they *have the right not to*.

Preventative medicine has proved to be the most effective way of reducing the incidence of fatal childhood diseases. As a result of the widespread practice of immunising young children in our society, *many lives have been saved and the diseases have been reduced to almost zero*. In previous centuries children died from ordinary illnesses such as influenza and tuberculosis and because few people had immunity, *the diseases spread easily*. Diseases such as dysentery were the result of poor hygiene but *these have long been eradicated since the arrival of good sanitation and clean water*. Nobody would suggest that *we should reverse this good practice now* because dysentery has been wiped out.

*Serious diseases such as polio and smallpox have also been eradicated* through national immunisation programmes. In consequence, children not immunised are far less at risk than they would otherwise be. Parents choosing not to immunise are *relying on the fact that the diseases have already been eradicated*. If the number of parents choosing not to immunise increased, there would be *a similar increase in the risk of the diseases returning*.

Immunisation is *not an issue like seatbelts which affect only the individual*. A decision not to immunise will have *widespread repercussions* for the whole of society and for this reason, I do *not* believe that individuals *have the right* to stand aside in this disease-free society. In my opinion immunisation should be obligatory.

### Follow up questions for discussions:

1. What or who is being evaluated in paragraph two?
2. What Attitude resource is used in the evaluation? Why?
3. How is the evaluation graded up or down in paragraph two? What Attitude resource is used?

---

<sup>24</sup> Jakeman, V. & McDowell, C. (2003) *Insight into IELTS extra: the Cambridge IELTS course*. Cambridge: Cambridge University Press. 185

4. Look at the first sentence of paragraph two. When the writer says, “preventative medicine has proved to be the most effective way of reducing the incidence of fatal childhood diseases”, what Engagement resource is employed? How strong is the opinion made? Does it open for other alternative opinions?
5. What or who is evaluated in paragraph three?
6. What Attitude resource is used in the evaluation? Why?
7. How is the evaluation graded up or down? What Attitude resource is used?
8. Look at the last sentence of paragraph three. When the writer says, “If the number of parents choosing not to immunise increased, there would be a similar increase in the risk of the diseases returning” what engagement resource is employed? How strong is the opinion made? Does it open for other alternative opinions?

## Session Four

### Part 1. Group work (trainer-trainees).

- A. Read the sample essay below and decide whether the following criteria can be identified in the essay.

<sup>25</sup>There is certainly a great deal of evidence to support the view that global warming is a result of the building-up of greenhouse gases caused by human activities. This essay will look briefly at the scientific debate surrounding its causes and, perhaps more importantly, consider what action can be taken to slow the process down.

Climate change has been happening since the late 19<sup>th</sup> century, but has increased significantly since then. The world’s oceans are steadily warming up and ice is melting at the north and south poles and mountain glaciers. Furthermore, scientists have demonstrated a clear link between human-made greenhouse gases and global warming. This finding clearly has serious implications for both human beings and animals alike. One outcome for humans will presumably be drought in areas where major water sources are fed by snow or glacier melt. For example, those who depend on river water are likely to be left without adequate water supplies during the summer. For animal and plant species such consequences are undoubtedly more serious. Animals like polar bears face extinction if their natural habitats continue to disappear. Some scientists, however, dispute the view that global warming is human-made. They claim that it can probably be explained by natural phenomena, such as changes in solar activity or variable changes in the climate.

It is the writer’s personal belief that the problem has largely been caused by human activity. Therefore, it is everybody’s responsibility to reduce greenhouse gas emissions, such as switching to technologies that produce little or none of them. People can obviously also make lifestyle changes such as reducing their number of car

---

<sup>25</sup> The essay is taken from *IELTS Preparation*. (2008, p. 172).

journeys and recycling their household rubbish. These small changes, surprisingly enough, can slow down the pace of global warming if people make the effort to do them. They can also increase the pace of global warming if they do not.

In conclusion, I agree with the statement that global warming is caused by human activity. It will change our way of life for the worse in the future if we do not make changes in the way we live now. If we continue to ignore the threat, we will surely regret it.

- A statement that immediately orientates the reader to the topic of the essay and how the writer will be approaching it. This might include some brief background to key concepts.
- The preview of position and main themes through which this position/argument will develop in the first paragraph.
- Development of main themes and their relationships in a series of well organised paragraphs which follow the order given in the first paragraph (are logically and clearly organised).
- Conclusion with overall interpretation, summarising the themes of the topic and highlighting their significance in light of the approach taken in the subject .

In SFL, which context do the criteria previously discussed belong to?

#### B. Has the sample text (essay) above met the criteria stated below?

- Causes of global warming and ways to slow it down
- Evidence and example to support argument
- Answer the question
- Formality (academic tone)
- Appraisal
- Textual cohesion

In SFL, which context do the criteria previously discussed belong to?

### Part 2. Pair work

Read the <sup>26</sup>sample essays below and decide whether the following criteria can be identified in the text (essay).

Nowadays, developing countries could be affected negatively by globalisation. There are some negative effects that should noticed carefully, as below.

Globalisation can create unhealthy competition in developing countries among local and foreign people. Globalisation makes developing countries widely open and exposed to the world. All access of information and people could be easily exchanged in every stage of society and business areas. This condition makes competition is more tight, local people that lack of knowledges, skills, informations, strategies, will fight harder to

---

<sup>26</sup> The essays are written by student participants identified as EAPs3 and EAPs5

survive in their business or career. Meanwhile, foreign people that usually come from developed nations will have more roles to control most strategic business or economy areas in developing countries. That definitely will bring huge benefits for them and their countries. Most of natural resources could be roled by them, and this condition would make the growth of developing countries become slower than they should be.

Besides that, globalisation also could threaten identity and characteristic of developing countries. Young generations could be easily affected by negative issues like violences, racism, pornography that have been exposed by almost every medias like newspaper, internet, magazine, television channels, movies, also video games. As result of that, principles of family that have been taught from every generation in developing countries start to fade out. That could be a great lost for the whole country.

To sum up, that are unhealthy competition among foreign and local people and could also threaten identity or characteristic of developing countries.

The term 'globalisation' refers to the process of allowing investment markets to work internationally due to the improvement in communication worldwide. Throughout the world, globalisation has affected the development of many countries, especially on developing countries. It can have a positive impact on developing countries, such as a greater chance of job availability for workers. Although it can have a positive effect on developing countries, it appears that globalisation has more negative impacts on developing countries.

One of the negative impacts of globalisation on developing countries is the exploitation of the workers done by the companies. An illustration of this, the sweatshop workers have to work overtime everyday and it is impossible for them to take a day-off eventhough they were sick. Instead of improving their wages, the companies keep paying their overtime workers underpaid.

Another negative impact of globalisation is that the health of people on developing countries are threatened by the negative effect of consuming junk food like McDonalds or KFC. Fast food chains that are spreading all over developing countries may affect the health of people who consume them too much. It is obviously will have a deep impact on the health of people consuming too much junk food.

To conclude, instead of having a positive impact, globalisation have more negative impacts on developing countries. The companies often exploit their underpaid workers and

the health of people are in dangers due to the existence of fast food chains on developing countries.

- A statement that immediately orientates the reader to the topic of the essay and how the writer will be approaching it. This might include some brief background to key concepts.
  - The preview of position and main themes through which this position/argument will develop in the first paragraph.
  - Development of main themes and their relationships in a series of well organised paragraphs which follow the order given in the first paragraph (are logically and clearly organised).
  - Conclusion with overall interpretation, summarising the themes of the topic and highlighting their significance in light of the approach taken in the subject.
- 
- Negative/positive impacts of globalization
  - Evidence and example to support argument
  - Answer the question
  - Formality (academic tone)
  - Appraisal
  - Textual cohesion

### Part 3. Individual work

**Rate the essay below by using the SFL-informed rubric given.**

Globalisation can have a bad impact on developing nations. Nowadays, there are a lot of company from a country make the other company on the other country hard because the price

There are some bad impact of globalisation. First, the economic of developing nations may disorder. The prices of the money on developing nations may decrease dramatically and all of the countries in the world may have the impact. Second, the citizen's life could be very hard because the prices of their daily use things may be very high. Especially for the poor people who usually only can buy food to eat. They could be very hard to eat because they do not have enough money to buy any food when prices rise very high.

The developing nations may have a serious problem in politics too when globalisation. Some of bad government may corrupt their country's money and go to other country to leave. It may make the country worse than before. Besides that, the citizen of the country may demonstrate because they cannot life well. The president may stress and the country may be poor, because the president could use the money of the

country to help the citizen. For example, the president may divide food to his people.

The impact of globalisation could make a big problem to the developing nations. It can destroyed a country because of the bad economic. A lot of citizen may move to the other country to have a better life. That's all may be bad impact which is happened because of globalisation.

## SFL-informed marking rubric<sup>1</sup>

### Marking Rubric

This SFL-informed marking rubric consists of 13 criteria which are used at the levels of genre, register, general academic essay features. These 13 criteria assess all the language resources that EAP-4 students use in their writing tasks. Each criterion is given a score of 0-3.

|   |                                                                                                |
|---|------------------------------------------------------------------------------------------------|
| 3 | Student's writing meets a criterion at the highest level expected as explicated in the rubric. |
| 2 | Quite strong evidence but not at the top standard                                              |
| 1 | There is evidence but it is weak                                                               |
| 0 | There is no evidence of the criterion                                                          |

| Genre: social purpose and stages developed in essays                                                                                                                                          |   |   |   |   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|---|---|
| Begins with a statement that immediately orientates the reader to the topic of the essay and how the writer will be approaching it. This might include some brief background to key concepts. | 0 | 1 | 2 | 3 |
| The first paragraph provides preview of position and main themes through which this position/argument will develop.                                                                           | 0 | 1 | 2 | 3 |
| Develops main themes and their relationships in a series of well organised paragraphs which follow the order given in the first paragraph are logically and clearly organised.                | 0 | 1 | 2 | 3 |
| Concludes with overall interpretation, summarising the themes of the topic and highlighting their significance in light of the approach taken in the subject.                                 | 0 | 1 | 2 | 3 |

| Register variables                                                                         |   |   |   |   |
|--------------------------------------------------------------------------------------------|---|---|---|---|
| Topic research and content (field)                                                         |   |   |   |   |
| Identifies and explores negative/positive impacts of globalisation                         | 0 | 1 | 2 | 3 |
| Uses information from prescribed text, uses evidence and examples to support argumentation | 0 | 1 | 2 | 3 |
| Answers the question                                                                       | 0 | 1 | 2 | 3 |

<sup>1</sup> Adapted from Rose, (2005), *Reading to learn-Book 2-Assessing writing*, p. 2.

| Communication roles (tenor)                                           |   |   |   |   |
|-----------------------------------------------------------------------|---|---|---|---|
| Formality (academic tone)                                             | 0 | 1 | 2 | 3 |
| Appraisal (word choices writers use to engage, persuade and evaluate) | 0 | 1 | 2 | 3 |

| Written features (mode)                                                                                                     |   |   |   |   |
|-----------------------------------------------------------------------------------------------------------------------------|---|---|---|---|
| Textual cohesion: sentence and paragraph construction and relationship, use of conjunctions, synonyms and pronoun reference | 0 | 1 | 2 | 3 |

| General academic essay features                            |            |   |   |   |
|------------------------------------------------------------|------------|---|---|---|
| Spelling and punctuation                                   | 0          | 1 | 2 | 3 |
| Grammatical accuracy (i.e. tenses, subject-verb agreement) | 0          | 1 | 2 | 3 |
| Adherence to word limit requirement                        | 0          | 1 | 2 | 3 |
| <b>Final grade (divided by 1.95)</b>                       | <b>/20</b> |   |   |   |
| Comments:                                                  |            |   |   |   |
|                                                            |            |   |   |   |
|                                                            |            |   |   |   |
|                                                            |            |   |   |   |

## Appendix F: Genre analysis of essay 8a and 10a

Genre analysis of essay 8a

| Stage             | Phase                | Text                                                                                                                                                                                                                                                                                      |
|-------------------|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Issue</b>      | <b>issue</b>         | The economies of the world are being integrated as a result of globalisation. Its impact on every country around the world has been a major discussion, but the topic which is frequently brought up is its impact on developing countries.                                               |
|                   | <b>preview</b>       | This argumentative essay will explain more about the impact of globalisation on developing countries.                                                                                                                                                                                     |
| <b>Side 1</b>     | <b>topic</b>         | Globalisation has given several advantages to developing countries.                                                                                                                                                                                                                       |
|                   | <b>elaboration-1</b> | The main advantage is how every country is now connected to each other.                                                                                                                                                                                                                   |
|                   |                      | Developing countries can gain access to the latest technology. Not only technology, but also education which is improving as a result of globalisation.                                                                                                                                   |
|                   | <b>elaboration-2</b> | The other advantages are employment of human resources due to the arrival of multinational companies in developing countries and competition in companies forcing them to lower price which benefits citizens in developing countries to get the best quality of products at a low price. |
| <b>Side 2</b>     | <b>topic</b>         | Despite all the positive impacts mentioned above, globalisation has affected developing countries in negative ways.                                                                                                                                                                       |
|                   | <b>elaboration-1</b> | The competition in multinational companies has led to the exploitation of labour.                                                                                                                                                                                                         |
|                   |                      | Wages are reduced and safety standards are ignored in order to produce cheap goods. Many local companies are taken over by foreign companies. Thus, many people in developing country are employed as workers, but the better positions are given to the people from rich countries.      |
|                   | <b>elaboration 2</b> | Multinational companies also affect the environment.                                                                                                                                                                                                                                      |
|                   |                      | Industrial waste which is thrown away in rivers and polluted gas which is freely disposed in the air might cause serious damage to citizens' health.                                                                                                                                      |
|                   | <b>elaboration 3</b> | Another negative impact {is} on local cultures in developing countries which are affected by foreign cultures through social media.                                                                                                                                                       |
| <b>Resolution</b> | <b>review</b>        | In conclusion, globalisation can cause positive and negative impacts on developing countries but the negative impacts are outweighing the positive impacts.                                                                                                                               |
|                   | <b>resolve</b>       | It causes the exploitation of natural resources and labour works. In addition, the benefit of globalisation is not universal. The rich are getting richer and the poor can get a better living but they are not going anywhere.                                                           |

### Genre analysis of Essay 10a

| Stage              | Phase                | Text                                                                                                                                                                                                                                                                                                                                |
|--------------------|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Thesis</b>      | <b>position</b>      | Today in developing countries can be impact by globalisation, the impact can be negative.                                                                                                                                                                                                                                           |
|                    | <b>preview</b>       | There are some negative impacts of globalisation in developing countries. The globalisation can impact of economy, culture of developing countries also differentiation of product.                                                                                                                                                 |
| <b>Argument 1</b>  | <b>topic</b>         | Globalisation can impact the economy of developing countries.                                                                                                                                                                                                                                                                       |
|                    | <b>elaboration 1</b> | With globalisation the economic of the country can be developed with cooperation with developed countries.                                                                                                                                                                                                                          |
|                    | <b>elaboration 2</b> | Globalisation also causes westernization. Some of people in developing countries are not using local brand. The sales of local brand is decrease, because of this many of domestic industrial are closed cause they cannot gain a profit, and the effect is many unemployed in the countries.                                       |
| <b>Argument 2</b>  | <b>topic</b>         | The other impact is there is no differentiation on product that represent the speciality from each country for example Coca Cola, and the develop country will gain the impact cause of globalisation.                                                                                                                              |
|                    | <b>elaboration 1</b> | Some of develop countries use employee of developing countries with cheaper wages to produce of their stuff. The standard of the world is all the same with no variety at all but they the material is from developing countries with cheapest price to buy. After the developed countries brand it, the price is become expensive. |
|                    | <b>elaboration 2</b> | Globalisation also affect the currency of the developing countries, cause developing countries import the branded products and not using local brands.                                                                                                                                                                              |
| <b>Restatement</b> | <b>review</b>        | Globalisation can impact the developing countries in negative way.                                                                                                                                                                                                                                                                  |
|                    | <b>restate</b>       | The developing countries will lose their culture because of globalisation. They are following the culture from mixed culture of developed countries and some of them also can follow the bad one. The local brands also disappear because of globalisation.                                                                         |