

AN ANALYSIS OF MODERATION DISCOURSE BETWEEN
ASSESSORS

CHUE SIU KUAN

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF MASTER
OF LINGUISTICS

FACULTY OF LANGUAGES AND LINGUISTICS
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

ABSTRACT

Moderation is a post-judgement process where two or more assessors discuss previously awarded marks with the intention of reaching a common score (Bloxham, Adie & Hughes, 2015; McNamara, 2000). Studies have found that a consensus is achieved through an exchange of rhetorical moves, which may be influenced by factors such as power or experience (Orr, 2007) however, few studies have focused on the ongoing talk process that occurs within the moderation session. This study examines the talk in paired moderation sessions and how they defend or justify the scores they have given to four students in a group oral interaction. The study employs the pragma-dialectical Ideal System for Resolution of Disputes (Van Eemeren et al, 1993) together with a “quality of argumentation” classification as analytical tools. In the moderation, the assessors justify and defend the scores they had given before coming to an agreed score. Findings indicate that agreement is achieved through concessions and compromises, and that each assessor justifies the score they had given through observations of the student’s performance within the group interaction. The findings may suggest implications for the role of moderation as a means of reliability in assessment.

ABSTRAK

Moderasi merupakan sebahagian daripada proses selepas penilaian di mana dua atau lebih penilai yang telah memberikan markah bertemu untuk membincang markah bagi tujuan mencapai markah yang dipersetujui semua pihak (Bloxham, Adie & Hughes, 2015; McNamara, 2000). Kajian-kajian lepas mendapati bahawa persetujuan dicapai melalui pertukaran retorik yang mungkin dipengaruhi faktor-faktor seperti kuasa atau pengalaman (Orr, 2007) tapi, tidak terdapatnya banyak kajian yang berfokus pada proses perbincangan yang berlaku dalam moderasi. Kajian ini meneliti perbincangan yang berlaku dalam sesi moderasi berpasangan serta bagaimana markah yang telah diberi kepada empat orang pelajar dalam interaksi lisan berkumpulan dipertahankan dan diwajarkan. Kajian ini menggunakan “Ideal System for Resolution of Disputes” dari pragma-dialectics (Van Eemeren et al, 1993) dan klasifikasi “quality of argumentation” sebagai alat-alat analisis. Dalam sesi moderasi, penilai-penilai mewajarkan dan mempertahankan markah yang mereka berikan dahulu sebelum sampai suatu markah yang boleh dipersetujukan. Dapatan kajian menunjukkan bahawa persetujuan dicapai melalui konsesi dan kompromi, serta setiap penilai mempertahankan markah yang mereka berikan melalui apa yang mereka perhatikan dari prestasi pelajar dalam interaksi berkumpulan tersebut. Dapatan kajian mungkin membayangkan implikasi pada peranan moderasi sebagai instrumen kebolehpercayaan dalam penilaian.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to those who have helped me in a multitude of ways in the process of completing this research. This work would not have been possible without them.

First, I would like to thank my supervisor, Dr. Teoh Mei Lin, for all her support and guidance. She lent me some books and guided me down the Argumentation theory path when I dithered between discourse analysis and speech acts. Additionally, she tirelessly read through version after version of analyses and drafts, and when I lost motivation, she spurred me on. I cannot express thanks enough for all her support, advice and guidance.

Secondly, and no less important, I would like to convey my thanks to the principal investigator of the UMRG project “Employers’ Assessment of Language Competence in Group Oral Interaction” (project number: RP014B-13SBS) for allowing me to use the data collected through the project. I would also like to thank the team members of the above named project whose questions challenged me to explore aspects and implications of argumentation and moderation I had never considered. Their advice, support and encouragement were invaluable.

Last but not least, I extend my thanks to friends and family who have seen me through this research. Thank you for your patience, your help, and your encouragement. Without everyone mentioned here, this research would not have been possible.

TABLE OF CONTENTS

	Page
TITLE PAGE	i
ORIGINAL LITERARY WORK DECLARATION FORM	ii
ABSTRACT	iii
ABSTRAK	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF EXCERPTS	x
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xiv
LIST OF APPENDICES	xv
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Background of the Study	1
1.3 Motivation for the study	4
1.4 Research Aim and Questions	5
1.5 Significance of the Study	6
1.6 Conclusion	8
CHAPTER 2 LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Pragma-Dialectics	9
2.3 Speech Acts	12

2.4	Speech Acts and the Four Meta Theoretical Principles	14
2.5	The Ideal System For Resolution of Disputes	14
2.5.1	Moves and Speech Acts	19
2.6	Studies Involving Pragma-Dialectics	22
2.7	Quality of Argumentation	24
2.8	Moderation	27
2.9	Studies Involving Moderation	29
2.10	Application of Argumentation to Moderation	33
2.11	Conclusion	35
CHAPTER 3 METHODOLOGY		36
3.1	Introduction	36
3.2	Theoretical Framework	36
3.3	Research Design	37
3.4	Data	37
3.5	Participants	33
3.6	Data Collection	39
3.6.1	Video Taped Group Oral Interaction	40
3.7	Data Analysis	40
3.8	Unit of Analysis	42
3.9	Reliability of the Study	45
3.10	Conclusion	46
CHAPTER 4 DATA ANALYSIS AND RESULTS		47
4.1	Introduction	47
4.2	Analysis on the Ideal System for Resolution of Disputes	47

4.2.1	Pair 1 (Assessors A1 and B1)	48
4.2.1.1	Discussion on Best Performer	49
4.2.1.2	Discussion on N4's TF Score	50
4.2.1.3	Discussion on N4's LC Score	50
4.2.1.4	Discussion on N4's CA Score	51
4.2.1.5	Further Analysis of Pair 1 Findings	52
4.2.1.6	Summary of Pair 1 Discussion Stages	55
4.2.2	Pair 2 (Assessors A2 and B2)	56
4.2.2.1	Discussion on Best Performer	57
4.2.2.2	Discussion on N2's TF Score	58
4.2.2.3	Discussion on N2's LC Score	58
4.2.2.4	Discussion on N2's CA Score	59
4.2.2.5	Further Analysis of Pair 2 Findings	60
4.2.2.6	Summary of Pair 2 Discussion Stages	64
4.2.3	Summary of Findings for Analysis Based on the Ideal System	64
4.3	Moves and Speech Acts Used in Argumentation Stage to Justify Scores Given	65
4.3.1	Pair 1 (Assessors A1 and B1)	66
4.3.1.1	Discussion on Best Performer	67
4.3.1.2	Discussion on N4's TF Score	69
4.3.1.3	Discussion on N4's LC Score	70
4.3.1.4	Discussion on N4's CA Score	70
4.3.1.5	Comparison of Speech Acts Used by A1 and B1	71
4.3.2	Pair 2 (Assessors A2 and B2)	73
4.3.3	Types of Moves and Speech Acts	77

4.3.3.1 Assertives	78
4.3.3.2 Commissives	79
4.3.3.3 Directives	80
4.4 Structure and Presentation of Justifications	81
4.4.1 Pair 1 (Assessors A1 and B1)	83
4.4.1.1 Discussion on Best Performer	84
4.4.1.2 Discussion on N4's TF Score	87
4.4.1.3 Discussion of N4's LC Score	89
4.4.1.4 Discussion of N4's CA Score	91
4.4.2 Pair 2 (Assessors A2 and B2)	96
4.4.3 Comparison between Pair 1 and Pair 2	101
4.5 Conclusion	102
CHAPTER 5 DISCUSSION OF FINDINGS AND CONCLUSION	103
5.1 Introduction	103
5.2 Ideal System for Resolution of Disputes	103
5.3 Moves and Speech Acts Used to Justify Scores Given	105
5.4 Structure and Presentation of Justifications	107
5.5 Overview of the Findings	108
5.6 Limitations and Avenues for Future Study	110
5.7 Conclusion	111
REFERENCES	114

LIST OF EXCERPTS

	Page
Excerpt 1.1 Pair 1 Transcript (Lines 115-123)	44
Excerpt 1.2 Pair 1 Transcript (Lines 43-49)	53
Excerpt 1.3 Pair 1 Transcript (Lines 96-107)	54
Excerpt 1.4 Pair 1 Transcript (Lines 56-58)	55
Excerpt 2.1 Pair 2 Transcript (Lines 1-32)	61
Excerpt 2.2 Pair 2 Transcript (Lines 202-216)	63
Excerpt 1.5 Pair 1 Transcript (Lines 12-23)	84

University of Malaya

LIST OF TABLES

		Page
Table 2.1	Distribution of Speech Act Types Across Functional Stages in Discussion	18
Table 3.1	Moderation Pairs and Details	38
Table 3.2	Participant Profiles for Moderating Assessors	39
Table 4.1	Discussion stages and their order in the discussion on best performer	49
Table 4.2	Discussion stages and their order in the discussion on N4's TF score	50
Table 4.3	Discussion stages and their order in the discussion on N4's LC score	51
Table 4.4	Discussion stages and their order in the discussion on N4's CA score	52
Table 4.5	Pair 1 discussion stages	56
Table 4.6	Discussion stages and their order in the discussion on best performer	57
Table 4.7	Discussion stages and their order in the discussion on N2's TF score	58
Table 4.8	Discussion stages and their order in the discussion on N2's LC score	59
Table 4.9	Discussion stages and their order in the discussion on N2's CA score	60
Table 4.10	Pair 2 discussion stages	64
Table 4.11	Summary of moves by speech act in argumentation stage 1 of discussion on best performer (Lines 12-37)	68
Table 4.12	Summary of moves by speech act in argumentation stage 2 of discussion on best performer (Lines 39-55)	69
Table 4.13	Summary of moves by speech act in argumentation stage 2 of discussion on N4's TF score (Lines 67-95) See Appendix H	159

Table 4.14	Summary of moves by speech act in argumentation stage of discussion on N4's LC score (Lines 115-123) See Appendix H	159
Table 4.15	Summary of moves by speech act in argumentation stage 1 of discussion on N4's CA score (Lines 163-230) See Appendix H	160
Table 4.16	Summary of moves by speech act in argumentation stage 2 of discussion on N4's CA score (Lines 265-278) See Appendix H	161
Table 4.17	Summary of speech acts by A1 and B1	72, 73
Table 4.18	Summary of moves by speech act in argumentation stage of discussion on N2's LC score (Lines 52-67) See Appendix I	162
Table 4.19	Summary of moves by speech act in argumentation stage 1 of discussion on N2's CA score (Lines 101-153) See Appendix I	162
Table 4.20	Summary of moves by speech act in argumentation stage 2 of discussion on N2's CA score (Line 161-176) See Appendix I	163
Table 4.21	Summary of moves by speech act in argumentation stage 3 of discussion on N2's CA score (Line 178-201) See Appendix I	164
Table 4.22	Summary of speech acts used by A2 and B2	77
Table 4.23	Comparison of the speech acts used by Pair 1 and Pair 2	77
Table 4.24	Categorization of assertives by moves made by Pair 1 and Pair 2	79
Table 4.25	Categorization of commissives by moves made by Pair 1 and Pair 2	80
Table 4.26	Categorization of directives by moves made by Pair 1 and Pair 2	81
Table 4.27	Quality of argumentation classification	83
Table 4.28	Summary of the Level of Arguments by Pair 1 (BP)	86
Table 4.29	Summary of the Level of Arguments by Pair 1 (TF)	89
Table 4.30	Summary of the Level of Arguments by Pair 1 (LC)	91
Table 4.31	Summary of the Level of Arguments by Pair 1 (CA)	96

Table 4.32	Summary of the Level of Arguments by Pair 2 (LC)	97
Table 4.33	Summary of the Level of Arguments by Pair 2 (CA)	100
Table 4.34	Comparison of arguments and Levels of Arguments of Pairs 1 and 2	101

University of Malaya

LIST OF ABBREVIATIONS

Abbreviation	Meaning
RQ	Research Question
N1	First student from left in Video Taped Group Oral Interaction
N2	Second student from left in Video Taped Group Oral Interaction
N3	Third student from left in Video Taped Group Oral Interaction
N4	Fourth student from left in Video Taped Group Oral Interaction
TF	Task Fulfillment (a section in the rating scale)
LC	Language Control, Vocabulary and Structures (a section in the rating scale)
CA	Communicative Ability, Fluency and Pronunciation Errors (a section in the rating scale)
BP	Best Performer (Used in tables presenting the findings of RQ3)

LIST OF APPENDICES

	Page	
Appendix A	Distribution of Speech Act Types Across Functional Stages in Discussion (Also Table 2.1)	118
Appendix B	Quality of Argumentation Classification (Also Table 4.27) Baartman et al's (2007) classification used as guideline	119
Appendix C	Rating Scale for Group Oral Interaction from course GTEE1109 Marksheet for Group Oral Interaction from course GTEE1109	120
Appendix D	Transcription Notation	123
Appendix E	Test Situation for Group Oral Interaction Task	124
Appendix F	Full Transcript Pair 1 (A1-B1)	125
Appendix G	Full Transcript Pair 2 (A2-B2)	145
Appendix H	Tables 4.13-4.16: Summary of Speech Acts and moves by Pair 1	159
Appendix I	Tables 4.18-4.21: Summary of Speech Acts and moves by Pair 2	162
Appendix J	Structure of Pair 2's (A2-B2) justifications	165
Appendix K	Analysis Summary of Pair 3 and Pair 4	170

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter will provide a brief introduction to the current study, by first providing a background to the research. The study involves an analysis of moderation talk that is obtained from assessment sessions in the context of a group oral interaction test. The talk which consists of the spoken discourse between pairs of assessors who were requested to award marks to four candidates in a simulated oral interaction task, after which each pair of assessors carried out the moderation which was audio-recorded. The moderation talk was then transcribed and the transcripts constitute the primary data used in the analysis for this study. This chapter provides an overview of some concepts such as moderation and discourse analysis, and also sets the backdrop for the rest of the dissertation. Also covered in this chapter will be the study's objective and its research questions, as well as a brief overview on how the study is carried out.

1.2 Background of the Study

Moderation is commonly considered as a “post-judgement process” (Bloxham, Hughes, & Adie, 2015, p. 2) where two or more assessors who have judged student performances come together to reach a consensus on the judgement or marks they had rendered (McNamara, 2000). While the goal of moderation is known, less is known of what exactly happens within the moderation process – just how do the assessors agree on this consensus? This appears to be an important consideration as moderation is a means of ensuring reliability given the subjectivity of the rating process (McNamara, 2000).

Research has been done on the various types of moderation (Bloxham, Hughes, & Adie, 2015), the underlying intentions of assessors faced with moderation (Adie, Lloyd, & Beutel, 2013), the importance of assessment criteria in the moderation process (Van der Schaaf, Baartman, & Prins, 2012), and even the theoretical approach of moderation nature has been debated (Orr, 2007). But there appears only to be one very brief study on the discourse within the moderation process (Orr, 2007), and this was to illustrate the theoretical approach underpinning moderation discourse.

As mentioned above, moderation is a means of ensuring reliability in assessment, making it a rather important part of the rating or assessment process. Assessment is an important part of teaching and learning, Boud (1995) indicates that conventionally, assessment is a means of measuring what the candidate has learned. It is thus acknowledged that assessment is an important part of the education process. However assessment is subjective as rater background and prior training may influence how one assesses and further training cannot completely eliminate this subjectivity (McNamara, 1996; 2000). This leads to a need to ensure reliability in assessment for the sake of fairness and objective testing, which calls for moderation as it is a means to ensure reliability.

While it appears that moderation and the role it plays in ensuring reliability is being debated by educators in countries such as UK and Australia (Bloxham, Hughes, & Adie, 2015; Adie, Lloyd, & Beutel, 2013), there still appears to be a long way to go to increase the reliability factor of moderation in practice, as studies have shown that it is a social activity that is highly subjective (Orr, 2007; Van der Schaaf, Baartman, & Prins, 2012). For despite the high hopes for objectivity and fairness, it was found that instead of being fair and positivist in nature as intended, moderation when practiced is a social, co-

constructed activity vulnerable to factors such as power, and other factors (Orr, 2007). Orr further cites previous studies where it appears that assessment and moderation by extension is a tool in which assessors use to either reward students for effort or to motivate the students to do better or to discourage them from complacency. During moderation, objectivity and criteria plays a big role, as shown by Van der Schaaf, Baartman, and Prins (2012) who argue that the use of criteria does aid in increasing the reliability of the assessment. However objectivity does sometimes fall by the wayside when assessors consider a particular grade the student 'deserves' in the pursuit of 'fairness' (Wyatt-Smith, Klenowski, & Gunn, 2010).

On the other hand, there have not been many studies done on talk between assessors in moderation sessions. In the literature one brief study by Orr (2007) illustrates the difference between the theoretical and highly objective positivist approach to moderation, in contrast with the subjective and co-constructed nature moderation, which is the usual practice. To my knowledge no studies have looked at talk during moderation, especially the process leading up to the determination of the agreed upon score. While the study will not be a discourse analysis in the strictest sense of its definition, it is a discourse-based analysis and will adhere to certain aspects of discourse analysis.

Discourse analysis, according to Jorgensen and Philips (2002), allows for the analysis of the use of language in various mediums to discern possible patterns of talk. What Jorgensen and Philips stress is that the analyst is limited to what is present within the analyzed discourse and that the role of the analyst is not to determine the truth of the statement or the intentions behind it. Similarly, this study will adopt the limitation of the analyst's role for the study seeks examine the talk process during moderation leading up to a consensus both agree on.

1.3 Motivation for the Study

The goal of moderation is known and accepted, and Orr (2007) has found that agreement is reached through an exchange of rhetorical moves. But few studies have examined the talk process, or the rhetorical moves through which assessors in a moderation reach a consensus on the scores they had awarded to students in an assessment. Given the nature of moderation where two or more assessors who have given differing scores are required to reach a consensus on said scores, and the relative lack of research on the ongoing talk process that occurs during moderation, this study sought then to examine the talk that assessors engage in as they reach a consensus on scores. The second aspect of the objective, how assessors justify their judgement, was incorporated into the objective to evaluate the how assessors justify the scores they had given since moderation plays an important assessment as a reliability measure.

The pragma-dialectic framework by Van Eemeren and Grootendorst (1984) was chosen as part of the scaffolding for the study based on the framework's role in the study of examining ongoing disputation and the resolution of said dispute. The argumentation framework appeared to closely parallel moderation where the moderating assessors discuss their different scores with the aim of agreeing on one final score. The Ideal System for Resolution of Disputes (Van Eemeren et al., 1993) was chosen as one of the analytical tools given its purpose as a guideline for analysis of the resolution of differences in opinions, and the obvious parallels between the ideal system and the general discursive patterns and outcome of moderation. This then led to the second research question of what moves and speech acts were used to justify scores given. Moves and speech acts were chosen as avenues of analysis as speech acts is an integral part of the pragma-dialectic

framework and form the backbone for the various verbal moves the participants of an argument engage in (e.g. advancing arguments, expressing standpoints etc.) in order to resolve the argument. The term moves in this question, and by extension the rest of the study, refers to the various argumentation moves suggested and listed in the Ideal System for Resolution of Disputes found in Van Eemeren et al. (1993). These moves function to further the ongoing discussion and aid in reaching a conclusion or an outcome acceptable to all parties (more on moves in Section 2.5.1).

However, it was found that logic and fallacies were not ideally suited to analysis aimed at examining how assessors justify their judgement as the study did not seek to find out the shortcomings of moderation in practice. A different analytical tool was then found in order to carry out analysis to fulfill the second part of the research objective. The Quality of Argumentation classification from Van der Schaaf, Baartman and Prins (2012) was adapted and operationalized as the second analytical tool as it is a proven means of evaluating arguments within the area of assessment research. As such, its use allowed this study to examine the justifications assessors put forth during the moderation sessions, and evaluate these justifications in terms of relevance and depth in order to examine, to a small extent, the role of moderation as a means of ensuring reliability in assessment.

1.4 Research Aim and Questions

This study will examine the talk that assessors engage in during the process of coming to a consensus on the scores awarded to four students in a group oral interaction, and how they justify this judgement. To achieve this objective, the study will attempt to answer the following research questions:

(a) In what way does the moderation discourse between assessors compare to Van Eemeren et al.'s (1993) pragma-dialectical Ideal System for Resolution of Disputes?

(b) In the argumentation stage(s), what moves and speech acts do the assessors use in justifying the score given?

(c) How are the justifications structured and presented?

1.5 Significance of the Study

The sections above have established the objective of the study, as well as provided a brief rationale behind choosing the analytical tools used in the study will be carried out. As was pointed out above, there have been several studies on moderation, but few on the moderation discourse itself. Orr (2007) found that the agreement on marks during moderation is reached through the exchange of rhetorical moves, but she did not divulge these rhetorical moves nor did she detail the talk process. This study will examine the moderation discourse using the pragma-dialectic framework which allows for both the examination of the talk process occurring in moderation as well as the evaluation of the rhetorical moves found in the moderation. It appears that there is not, to date, any studies on moderation which uses the pragma-dialectic framework as an analytical tool, hence one reason for the current study. Additionally, the use of the Quality of Argumentation Classification allows for the evaluation of what justifications the assessors put forth for the scores they had given, and may shed light on how assessors of group oral interaction assign scores for a student and their ability to justify their judgment.

More importantly, given the goals of moderation as a means of ensuring reliability in assessment, particularly the assessment of group oral interaction which is simultaneously fleeting in nature, highly complex and demanding on the assessor, it is imperative to determine the efficacy of moderation in this setting. Due to the subjective nature of assessment (McNamara, 2000), the varying perceptions of the role of moderation (Adie, Lloyd, & Beutel, 2013), and the many influencing factors of moderation (Orr, 2007; Van der Schaaf, Baartman, & Prins, 2012), it appears that moderation in practice may not fulfill its intended role. Therefore, this study aims to examine the talk occurring in moderation and the reasoning behind the generated talk in hopes of identifying what advantages and disadvantages exists in the practice of moderation in this public higher education institution.

It is hoped that the study will provide some insight on how assessors view and practise moderation, as well as whether or not moderation is the reliability enforcement tool it was intended to be. Part of the reason for identifying what moderation practices are utilized is also, to some extent, to aid in improving the application of moderation as a means of ensuring reliability in assessment. And as the course and test used in this study is an in-house test, it seems even more pressing that the moderation discourse be analyzed, and any shortcomings identified may hopefully be rectified to ensure that the assessment is fair and reliable.

1.6 Conclusion

This chapter has provided a brief background of the study focusing on moderation, the motivation behind the study, the aims and research questions, as well as the significance of the study. The next chapter will discuss more thoroughly the two argumentation based analytical tools used in the study: pragma-dialectics and the Quality of Argumentation classification, the role and scope of moderation, as well as provide some supporting literature regarding both the analytical tools and the subject of moderation.

University of Malaya

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter will cover some of the related literature in the area of the current study which focuses on the discourse in moderation sessions. Undergirding the study is the theory of argumentation and speech acts. Some previous studies done in this area are also included.

2.2 Pragma-Dialectics

Pragma-dialectics is an argumentation theory put forth by Van Eemeren and Grootendorst in 1984 for the analysis of arguments in actual practice, and is the study of ongoing disputation and the logic or reason applied in arguing the dispute (Van Eemeren & Grootendorst, 1984; Van Eemeren et al., 1993). In other words, this theory examines discussions in which there exists differences in opinion and how the parties involved in the discussion utilize logic and reason to resolve these differences and reach a conclusion satisfactory to all parties involved. While the words argumentation and dispute typically bring to mind contentious, confrontational debates; that is not the case in argumentation theory. According to Van Eemeren and Grootendorst. (2004, p. 1), argumentation is a “verbal, social and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by advancing a constellation of propositions justifying or refuting the proposition expressed in the standpoint”. Simply put, argumentation is a discussion whereby a set of reasons, explanations, elaborations etc. are presented to support or reject a particular stance adopted by one party in the discussion, and these reasons or arguments are aimed at convincing or swaying parties who doubt said stance.

A standpoint is a specific point of view regarding the issue in question, and must be defended with argumentation (Van Eemeren & Grootendorst, 2004), it is also described as a position taken regarding “a certain proposition in which a certain property or quality is ascribed to the persons or the things referred to” (Van Eemeren, Grootendorst, Snoeck Henkemans, 2002, p. 5). A standpoint may be expressed either explicitly or implicitly, in the former, the point of view or position has been clearly stated whereas the point of view in the latter has to be surmised from what is said. Argumentation differentiates between positive and negative standpoints, in the former, the party has adopted a position in support of the proposition from which the difference of opinion arises and in the latter the position is one that opposes the proposition. A positive standpoint is a position in support of the proposition while the negative standpoint is one which opposes the proposition. Van Eemeren, Grootendorst and Snoeck Henkemans (2002) differentiates between expressing doubt and expressing a negative standpoint. Doubt is often implicit, and unlike a negative standpoint, the speaker does not undertake a commitment to defend it.

According to Van Eemeren, Grootendorst and Snoeck Henkemans (2002), a dispute is merely a difference of opinion regarding a particular issue or proposition, the difference may well be simple, like when a standpoint encounters doubt. The authors identify four types of differences of opinions, these being single non-mixed, single mixed, multiple non-mixed and multiple mixed. The single non-mixed occurs when there is only a single difference of opinion whereby a standpoint meets with doubt. The single mixed, there is still only one difference of opinion but opposing standpoint or positions regarding the issue in question. In a multiple non-mixed, there are multiple differences or multiple issues in contention, but no opposing standpoints, merely expression of doubt. A multiple mixed happens when there are opposing standpoints for the issues in contention.

According to Van Eemeren et al. (1993), pragma-dialectics is an integrated theory capable of examining both the descriptive and normative aspects of an argument, in which the descriptive aspect encompasses the ongoing disputation process while the normative aspect the logic and reasonableness of the argument. The theory therefore is able to look at the verbal interaction the participants of the argument engage in, as well as the soundness of the arguments they put forth. This feat is made possible by the four meta theoretical principles put forth by the authors, these being externalization, socialization, functionalization, and dialectification which are in turn based on speech acts. These four meta theoretical principles are, for lack of a better term, the backbone in the study of argumentation, and are the tools through which the descriptive and normative approaches to studying are integrated. Van Eemeren et al. defines externalization as involving the expression of the difference of opinion and to acknowledge that there is a difference. This expression and acknowledgement is crucial as argumentation typically involves two or more parties who hold differing views and if the differences are not externalized, there is no basis for argumentation. Externalization also, to some extent, reveal the reasoning patterns and processes of the parties involved in the argumentation as they justify their views. The authors indicate that externalization is quite individual centric as it regards mainly the views and arguments of each party without considering the interaction occurring within the argument, as opposed to socialization. Socialization, according to Van Eemeren et al. is the view of argumentation involving two or more individuals trying to solve a problem or resolve a difference and the presence of two opposing parties is crucial to this principle. However, argumentation is not just two parties with differing opinions who try to justify their stances, one crucial aspect of argumentation is the interaction these two parties engage in as they argue their points, attempt to convince the opposing party, and try to reach a reasonable agreement. The authors further state that

socialization demands that there be the participant roles of protagonist and antagonist within the argument where both protagonist and antagonist work together to build the arguments with the aim of reaching an agreement. Functionalization, as Van Eemeren et al. explain, takes the view that there must be a purpose to the argumentation, where the individuals or parties involved endeavor to resolve their difference through reason. However, the authors add that the goal of resolving differences is only one part of functionalization for it also takes into account the ongoing process of resolving said differences and the quality and effectiveness of the arguments put forth in attempting to resolve the difference. And all of this hinges on whether the participants of the argument had adhered to the structure of argumentation which then leads to the next meta theoretical principle. Van Eemeren et al. put forth dialectification as the factor that brings together the three meta theoretical principles mentioned here in necessitating that argumentation possesses a set of rules or procedure to the purpose of resolving the difference. The authors identify dialectification as a set of rules or procedure that is manifested in the speech acts that must be performed, and performed according to a particular structure aimed at resolving the expressed difference. In other words, dialectification imposes organization to the argumentation, allowing the argumentation to be viewed as a critical discussion, and idealizes the argumentative discussion.

2.3 Speech Acts

As is mentioned above, speech acts are an integral part of the pragma-dialectic theory, providing the foundation for Van Eemeren and Grootendorst's argumentation theory. This is made possible by the nature of speech acts given that we do things with words is the message behind Speech Act Theory. Austin introduced three facets of a speech act, the locutionary, illocutionary and perlocutionary acts (Huang, 2007; Yule, 1996), and

differentiated between direct and indirect speech acts (Robinson, 2006). Austin also stipulated felicity conditions required for a valid speech act which Searle further developed by putting forth the categories of propositional content, preparatory condition, sincerity condition, and essential condition (Finegan, 2008; Fromkin, Rodger, & Hyams, 2011).

According to Huang (2007) and Yule (1996), the propositional content refers to verbal content of the speech act. The preparatory conditions are the circumstances behind the act where the act will likely not occur on its own, and that the interlocutors involved are capable of fulfilling the act. The sincerity condition demands that there is genuine intention behind the utterance. And the essential condition is that the intention behind the act is recognized, and that there is a change in the state of obligation in the person responsible in carrying out the act. These felicity conditions are required for a successful speech act and the conditions for a request may differ from those for a command as the context and circumstances for the two different types of acts do differ. Thus, Austin classified speech acts into different categories, and many revisions of his taxonomy were made, the most well-known being Searle's (Robinson, 2006).

Searle classified speech acts by grouping them into five distinct types (Cutting, 2008), these being declarations, representatives, commissives, directives and expressives. Declarations encompasses ritual pronouncements and so on, representatives (also known as assertives) function to convey information, commissives commit the speaker to a course of action, directives are the means of making the recipient do what the speaker wants them to do, and expressives are used to express the speaker's feelings (Robinson, 2006).

2.4 Speech Acts and the Four Meta Theoretical Principles

According to Van Eemeren et al. (1993), a large part of argumentation involves convincing another party and being convinced of the acceptability of a standpoint. In terms of externalization, this act of “being convinced”, they indicate, can be viewed as committing oneself to the conditions stipulated by the accepted speech act, and in voicing acceptance, one verbally commits oneself. In socialization, the stances and arguments put forth by the various parties, and the way the parties react in their opposing roles can be viewed as speech acts the parties are accountable to. In functionalization, reasonable arguments are put forth with the aim of resolving the difference, hence the speech act underscores the organization of the arguments and the way in which these arguments contribute to the resolution. Dialectification comprises a set of rules and speech acts idealized in the form of a critical discussion, this, they state, is comparable to the way in which Searle’s felicity conditions are “idealized model of an act type” (p. 16).

2.5 The Ideal System for Resolution of Disputes

In line with the meta theoretical principle of dialectification, Van Eemeren et al. (1993) put forth an Ideal System for Resolution of Disputes based on the rules for argumentation. The system hinges on the condition that a resolution is reached only when all the parties in the discussion are in agreement on the acceptability of the standpoints or vice versa. This means that one party must accept the arguments and standpoints of the other and in turn withdraw his standpoint because his arguments does not survive criticisms.

Van Eemeren and Grootendorst (2004) differentiates between resolution and settlement. A settlement occurs when the difference in opinion is put to an end through means other

than argumentation. Examples of settlement include a vote or an unbiased third party arbitrator. They stress that resolution is achieved only when all parties agree on the acceptability of the standpoints through thorough argumentation.

According to Van Eemeren and Grootendorst (2004), this ideal system serves both heuristic and critical functions in analyzing and evaluating a discussion. The system itself is a guideline for the analysis of the discussion, as it could aid an analyst or researcher in identifying and understanding all aspects relevant to a critical evaluation, hence its heuristic function. In terms of its critical function, the system supplies a set of rules or standards through which helps to determine how a discussion deviates, in any way, from the most conducive path towards resolution. Thus, an analyst can, through this ideal system, examine why and how some discussions fail to be resolved.

As mentioned earlier, there are four discussion stages within this ideal system: 1) the confrontation stage, 2) the opening stage, 3) the argumentation stage, and 4) the concluding stage. Each stage consists of verbal moves with functions particular to the respective stages. While it is not necessary that all four stages be explicit in a discussion, or occur in the correct order, a resolution is reached only when the parties involved have progressed reasonably through each stage, be it explicitly or implicitly (Van Eemeren et al., 1993).

In the confrontation stage, differences in opinions surface as it becomes apparent that a standpoint encounters doubt or contradiction, resulting in non-mixed or mixed difference of opinion respectively (Van Eemeren, Grootendorst & Snoeck Henkemans, 2002). The doubt or contradiction may or may not be explicitly expressed, and may oftentimes be implicit.

In the opening stage, the parties involved in the discussion attempt to establish what common ground or similarities they share in order to determine if there is a possibility of resolution. It is also in this stage that participant roles of protagonist and antagonist are distributed, whereby the protagonist will defend the standpoint in contention and where the antagonist will respond critically to the standpoint and the arguments put forth by the protagonist (Van Eemeren et al., 1993). In the case of a mixed difference of opinion, the parties may hold the roles of both protagonist and antagonist with respect to their own and the opposing party's standpoint. In practice, the participants in the discussion will establish participant roles and determine if there is enough shared ground for a fruitful discussion.

In the argumentation stage, the protagonist advances arguments to defend his standpoint with the intention of disproving the antagonist's doubts or rebutting any counter arguments from the antagonist (Van Eemeren & Grootendorst, 2004). The antagonist will then determine if he finds the argument acceptable, should he not accept either a part of or the entirety of the argument, he may request further argumentation or refute what has already been put forth, leading to further argumentation from the protagonist. It is crucial that both protagonist and antagonist be critical in their arguments and reactions here to maximize the possibility of a resolution.

In the concluding stage, the parties establish the result of the arguments put forth in the argumentation stage. The difference in opinion is reached only when both parties agree that the protagonist's standpoint is acceptable, at which the antagonist must retract his doubt, or conversely, if both agree that the protagonist's standpoint is not acceptable, then the protagonist must retract it (Van Eemeren, Grootendorst & Snoeck Henkemans, 2002).

Van Eemeren and Grootendorst (2004) identify the types of speech acts that contribute to the resolution of disputes and the roles they play. According to them, three types of speech acts: assertives, commissives, and directives play major roles in the resolution of disputes. Declaratives is present only in the form of usage declaratives, which function to aid in defining, specifying, amplifying etc. Expressives, they say, do not have a direct role, since they do not commit the speaker, but may influence the outcome of the resolution process.

Van Eemeren and Grootendorst (2004) describe assertives as speech acts through which a proposition is asserted and views this as a means through which the speaker commits himself to the statement given. Assertives are usually utilized in expressing a standpoint, advancing arguments, and retracting or upholding a standpoint. They list assertions, statements, claims, denials etc. as assertives through which arguments may be advanced, and indicate that there are various degrees of assertives by comparing a strong assertion against the weaker supposition. Commissives are described as speech acts where the speaker undertakes or does not undertake a responsibility such as accepting, and generally committing themselves to an action or a stance. These include “accepting or not accepting a standpoint, accepting a challenge to defend a standpoint, deciding to start a discussion, agreeing to assume the role of protagonist or antagonist, agreeing to the discussion rules, accepting or not accepting argumentations, and deciding to start a new discussion” (p. 65)

Van Eemeren and Grootendorst indicate that the commitments undertaken may not necessarily be to the advantage of the hearer. Directives involve the speaker’s attempt at getting the hearer to do something. Directives used in critical discussions involve challenges and requests in which one party challenges the other’s standpoint, or requests explanations, definitions or further arguments. Van Eemeren and Grootendorst stress that not all directives are used in critical discussions, orders and prohibitions have no place in

such discussions. They further state that the protagonist can only be challenged to defend a standpoint and not to a fight.

Given the fact that each type of speech act has its respective roles in a critical discussion, Van Eemeren et al. (1993) proposed a model or a system in which the speech acts are distributed over the different stages of the critical discussion. In this model/system, they spell out the moves performed in each stage and their corresponding speech acts (as seen in Table 2.1, also in Appendix A).

Table 2.1 Distribution of Speech Act Types across Functional Stages in Discussion

Stage in discussion	Speech act type
Confrontation	
1.1	Expressing standpoint (assertive)
1.2	Accepting or not accepting standpoint (commissive)
Opening	
2.1	Challenging to defend standpoint (directive)
2.2	Accepting challenge to defend standpoint (commissive)
2.3	Deciding to start discussion; agreeing on discussion rules (commissive)
Argumentation	
3.1	Advancing argumentation (assertive)
3.2	Accepting or not accepting argumentation (commissive)
3.3	Requesting further argumentation (directive)
3.4	Advancing further argumentation (assertive)
Concluding	
4.1	Establishing result (assertive)
4.2	Accepting or withholding acceptance of standpoint (commissive)
4.3	Upholding or retracing standpoint (assertive)
(any stage)	
5.1	Requesting usage declarative (directive)
5.2	Defining, precizing, amplifying, etc. (usage declarative)

(Source: van Eemeren et al., 1993, p. 31)

2.5.1 Moves and Speech Acts

As has been mentioned, Van Eemeren and Grootendorst. (2004) define argumentation as a “verbal, social and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint by advancing a constellation of propositions justifying or refuting the proposition expressed in the standpoint” (p. 1). As such, in order to proceed with the argument and convince the opposing party, the participants employ a series of verbal moves, and these moves are specific to the respective discussion stages of confrontation stage, opening stage, argumentation stage, and concluding stage. Additionally, is governed by a particular speech act characterizing the nature of the move (Van Eemeren et al., 1993).

As Van Eemeren, Grootendorst and Snoeck Henkemans (2002) state, the confrontation stage is where differences in opinions surface. Table 2.1 indicate two moves characteristic to the confrontation stage, these being *1.1 Expressing standpoint* and *1.2 Accepting or not accepting standpoint*. *Expressing standpoint* is where one or all parties involved in the discussion state their opinion or present their stance. In the context of this study, this move is identified as when the moderators state the score they had given to the student during assessment. Once a standpoint has been expressed, the parties involved in the discussion each have the option of accepting or not accepting said standpoint (move 1.2 in Table 2.1). Should the standpoint be accepted, then the discussion ends as the difference in opinion has been resolved, however if the other parties choose not to accept said standpoint, the discussion then proceeds to the opening, argumentation and concluding stages. In the context of this study, non-acceptance of standpoint is seen when the second moderating assessor does not explicitly state their agreement to the marks the

first moderating assessor has given the student, or if they state that they had given that student a different score.

Once the difference in opinion(s) has been identified and established, the participants of the argument would, ideally, proceed to the opening stage where they try to establish common ground in order to determine possibility of resolution (Van Eemeren et al., 1993). This stage, as seen in Table 2.1, possesses three characteristic moves: 2.1 *Challenging to defend standpoint*, 2.2 *Accepting challenge to defend standpoint*, and 2.3 *Deciding to start discussion*. The move *Challenging to defend standpoint* occurs when a standpoint is not accepted (move 1.2) and the opposing party requests or demands reasons for this particular standpoint. In the context of this study, this is manifested when a moderating assessor asks their counterpart for the reasons why they had awarded a certain score to the student. In the move *Accepting challenge to defend standpoint* the challenged party takes on the role of protagonist and is obliged to advance arguments for their standpoint. The move *Deciding to start discussion* encompasses determining the discussion rules for the upcoming argument. This move, the authors indicate, is often implicit. This is also true in the context of this study as the role and scope of moderation limits the arguments each party may or may not put forth to what they have assessed from the assessment. It is possible that the rating scale of the assessment may be adopted as part of the discussion rules.

After assigning participant roles and determining discussion rules, the participants enter the argumentation stage where arguments are put forth to defend the standpoint in order to silence the antagonist's doubts (Van Eemeren & Grootendorst, 2004). This stage is characterized by the following moves: 3.1 *Advancing argumentation*, 3.2 *Accepting or not accepting argumentation*, 3.3 *Requesting further argumentation*, and 3.4 *Advancing*

further argumentation. Every move in this stage revolves around argumentation, either in advancing arguments, accepting or not accepting them, and the request and advancement of further argumentation. In the context of this study, argumentation should consist of reasons why the assessor had awarded a certain score to the student. In this stage, the opposing party may accept the arguments if he or she finds them acceptable, or keep requesting further argumentation until they are satisfied. The protagonist, or defending assessor in this case, is obliged to advance arguments when requested until the antagonist is convinced.

This then leads to the concluding stage, where the result of the argument is established based on the arguments put forth, and one party must relinquish their standpoint and accept that of the other party's should their own standpoint be found unacceptable (Van Eemeren., Grootendorst & Snoeck Henkemans, 2002). A standpoint is not acceptable if found unable to stand up to criticism (Van Eemeren et al., 1993). The three moves of this stage are *4.1 Establishing result*, *4.2 Accepting or withholding acceptance of standpoint*, and *4.3 Upholding or retracing standpoint*. The move *Establishing result* occurs when one or both parties determine if the arguments advanced in support of the standpoint were acceptable. This decision then leads to *Accepting or withholding acceptance of standpoint*, following which said standpoint may be upheld or must be retracted (move 4.3 in Table 2.1). In the context of this study, the successful defense and acceptance of a standpoint or the awarded score would require, ideally, that the defending party put forth argumentation the opposing party finds acceptable, following which the defended score is accepted and upheld, which then means that any other standpoint or score put forth within the discussion must be retracted, leaving only one final score, the defended one.

As seen in Table 2.1, each of these moves fall under different types of speech acts, for instance *Expressing standpoint*, *Advancing argumentation*, and *Establishing result* are assertives as the speaker asserts a proposition and commits themselves to it (Van Eemeren & Grootendorst, 2004). Accepting or not accepting a standpoint or argument (moves 1.2, 3.2 and 4.2 in Table 2.1) are labeled as commissives as these moves require the hearer to undertake the responsibility of accepting or not accepting and thereby committing or not committing to a stance, as Van Eemeren and Grootendorst explain. The moves *Challenging to defend standpoint* and *Requesting further argumentation*, Van Eemeren et al. (1993) have identified as directives as these moves require or demand a particular action from the hearer.

Together, these moves and speech acts form the various stages of the Ideal System for Resolution of Disputes and provide a functional guide for the analysis and evaluation of arguments and critical discussions. Van Eemeren et al. (1993) stress, however, that the ideal system plays the role of a guide to the analysis and evaluation of arguments, and that it is, in no way, a hard and fast model all arguments must conform to. It is, as its name suggests, a system detailing how an ideal resolution of disputes may play out.

2.6 Studies Involving Pragma-Dialectics

Lewinski (2010) examined online political forums as an argumentation context. Lewinski had attempted to examine this argumentation context with two other argumentation frameworks before settling on the pragma-dialectical theory for argumentation. Lewinski attributed the tendency towards critical reaction in the online forums to the pseudonymous environment within that context, and some patterns of these critical reactions were then

analyzed as strategic maneuvering. Lewinski found that the pragma-dialectical theory was suited to the critical reactions and strategic maneuvering present within the online political forum context.

Pedersen (2010, p. 1) examined the feasibility of “reasonable non-agreement” despite conforming to “pragma-dialectic rules for a critical discussion”. The author analyzed the claim that such a scenario is possible put forth by Van Eemeren et al. (1993) and compared it with various other argumentation rules. While it seems that a “reasonable non-agreement” is possible, the author advises a need for more extensive research.

While it may seem that most studies employing pragma-dialectics as a framework appear to focus on mediated disputes (Aakhus, 2003) and rhetoric (Kaptein, 2005; Zarefsky, 2006), pragma-dialectics can and has been used to analyze non-confrontational discourse such as problem solving discussions or negotiations.

Thirlwall (2008) investigated problem solving in negotiations. The study is both theoretical and prototypical as he explored a pragmatic approach towards problem solving strategies in negotiation through speech acts and pragma-dialectics. He analyzed the functions of speech acts within the commitment aspects of negotiation and how these commitments may be challenged and ‘repaired’. He identified ‘repair’ as a move incorporating the criticisms aimed at negotiation commitments. He then reconstructed the argumentation structure for some proposals found Mohammed’s (2005) data, then demonstrated the relationship of elements within the proposal and sketched a strategy for ‘repairing’ or “renegotiating broken agreements”. He found that a discourse based pragmatic approach to negotiation offers a systematic means of reconstructing, analyzing, and evaluating the negotiation.

Mohammed (2006) explored the advantages of investigating negotiations through pragma-dialectics. She argues that pragma-dialectics allows for the analysis of the ongoing process of interaction within the negotiation without being colored by psychological speculations and provides a means for critically evaluating the rationality of negotiation encounters as it allows for the reconstruction of the negotiation discourse. She further argues that negotiation is an argumentative activity since negotiation is the process of resolving a conflict of interest and each party attempts, as far as possible, to satisfy its own interests, then demonstrates the argumentation structure of a negotiation excerpt. She concluded that in reconstructing the negotiation, it is possible to identify why and how it may have failed.

Van Rees (2003) demonstrated the advantages of approaching the analysis and subsequent improvement of problem solving discussion using pragma-dialectics. She lamented the inadequacy of previous approaches in enhancing the quality of problem solving discussions, indicating there was little analysis of the interaction within the discussion, and the failure of these approaches in evaluating the quality of the discussions. She identified and demonstrated the ability of Van Eemeren and Grootendorst's pragma-dialectic theory, in particular the model of critical discussion, in both analyzing the interaction within a discussion in a medical institution and evaluating the quality of said problem solving discussions.

2.7 Quality of Argumentation

It can be seen above that the pragma-dialectic theory is a rather comprehensive theory which allows for the examination of both the ongoing argumentation process and its

reasoning. This, as noted by Van Rees (2003), is highly different from many other theories involving argumentation which usually focus only on either one of these aspects. And while the ideal system very ably provides a guide for the argumentation process complete with basic moves, it does not allow for the evaluation of the reasoning of the argumentation. In pragma-dialectics, the evaluation of the arguments is achieved through the study of logic and fallacies where fallacies occur when the discussion rules have been violated (Van Eemeren, Grootendorst & Snoeck Henkemans, 2002). And while rhetoric and strategic maneuvering has been integrated into pragma-dialectics (Van Eemeren & Houtlosser, 2002; 2006), it does not appear to be a suitable fit with the current objective. Although the study of fallacies committed by assessors during a moderation session would likely be enlightening, it would likely detract from the objective of this study which focuses on the moderation process and how the assessors justify their scores. This is because the study of fallacies would redirect the study to the argumentative faux pas or fouls committed by the moderating assessors, taking the study away from the structuring of arguments to justify their scores.

Instead, another model or framework is required to achieve the score justification aspect of the objective. Toulmin's (1958) model for argumentation is often referred to for argument structure as it allows for the analysis of the rhetorical moves in argumentative discussions. Toulmin's model has been successfully used in the classroom context, in the fields of composition studies (Bizup, 2009) and classroom scientific discourse (Driver, Newton, & Osbourne, 2000), and even in television political interviews (Lauerbach, 2007). However, it does not appear to allow for the evaluation of argument quality which would be a consideration in how assessors justify scores during moderation. Hence, studies involving assessment and moderation were referred to for a suitable argument evaluation model. Van der Schaaf, Baartman, and Prins (2012) were found to have used

a quality of argumentation classification by Clark and Sampson (2007) as they explored the teacher judgement process.

Clark and Sampson (2007) explain their reason for looking at argument quality, stating that Toulmin's model of argumentation allows for an examination of argument structure in how each rhetorical exchange relates to others within that same argument topic. However, these rhetorical exchanges may be undermined by their content should a certain move bear incorrect or irrelevant information. This, they argue, is reason for examining argument quality and had based their classification on Erduran, Simon, and Osborne (2004). The classification comprises 5 levels of argument quality ranging from no support in Level 1 to highly coordinated support based on ideas and evidence in Level 5 (refer Table 4.27 or Appendix B for operational definition). These varying levels of argument quality is based on the relevance of the information or evidence used in supporting an argument, as well as what connections have been drawn between the evidence and certain concepts or ideas.

Van der Schaaf, Baartman, and Prins (2012) successfully used this classification to demonstrate the importance of the use of criteria during assessment and moderation. They found that teacher judgement were found to be of higher quality when the criteria was in use. Furthermore, Van der Schaaf, Baartman, and Prins (2012) cited successful use of the classification in Baartman et al.'s (2007) study on the quality of assessment programs (refer Appendix B for Baartman et al.'s classification).

2.8 Moderation

Moderation has long been considered a “post-judgement process” (Bloxham, Hughes, & Adie, 2015, p. 2), it has been defined as “the process of reconciling or reducing differences in the judgments and standards used by different raters within a rating procedure” (McNamara, 2000, p. 134), and is “often a form of second or double marking of summative assessment” (Bloxham, 2009, p. 212). Adie, Lloyd, and Beutel (2013) identifies two types of moderation: internal moderation and external moderation; the former a test of assessment quality within the institution and the latter as a means to synchronize with national standards. It appears, recently, many studies on moderation revolve around accountability and quality assurance (Johnson, 2015; Kohoutek, 2014).

According to Bloxham, Hughes, and Adie (2015) recent studies point to moderation being a cyclic process, where various moderation practices have been identified and associated with different stages of the assessment process, each fulfilling their own respective purposes. The practices contribute to the assessment stages of design, calibration, judgements, external moderation/comparison, and the monitoring of evaluation; making moderation a form of verification and quality assurance in assessment, where it functions as a means of “agreeing, assuring and checking standards” (p.1). The authors provide some of the purposes of moderation in the different stages, such as ensuring the quality of the course and its tasks (design), creating a shared understanding of specifications, requirements, standards, and criteria for said tasks (calibration), insuring judgement quality through criteria adherence, credibility of evidence, shared standards, and consistency of judgement (judgement), comparing and benchmarking standards (external validation), and ascertaining the overall quality of assessment with attention given to the individual stages and their respective components (monitoring evaluation). They further

indicate that these various goals of moderation are achieved through activities such as peer scrutiny, socialization, workshops, collaborative marking, consensus moderation discussions, external examining, professional accreditation process etc. As a result of the increasingly extensive role of moderation in ensuring reliability within assessment, many countries and examination standards appearing to be jumping on the moderation bandwagon given this age of concern over reliability, consistency of judgements, and quality in assessment.

Orr (2007) contrasts the positivist and poststructuralist perspectives in assessment research. According to her, the positivist perspective focuses on objectivity or a right or ideal score, and research on “techno-rationalism, focus on the concept of transparency, the use of scientific language, emphasis on measurement, and an emphasis on freestanding standards” (p. 646) are indicative of this approach. She further explained that transparency in assessment, the validity of the means used to measure performance, and the existence, use of standards in assessment and ultimately fairness in assessment are the backbone of the positivist approach. The poststructuralist perspective on the other hand, view assessment as a co-constructed activity which is influenced by relations of power, in which judgment of performance is interpretative and thus subjective. She identified moderation as possibly falling within the poststructuralist perspective due to the contextualized nature of the discourse within moderation.

While moderation appears to be a sound means of ensuring reliability and accountability, this practice appears to have its disadvantages as well. Adie, Lloyd, and Beutel (2013) cite numerous shortcomings related to moderation, indicating that while the purpose of moderation is to support the practices of teaching and learning and to ensure the accountability of judgements, studies have shown that assessors have trouble balancing

these two processes, inadvertently focusing on one at the expense of the other. They further pointed out that while the implementation of criteria and moderation appears to be advantageous in increasing consistency in marking, the criteria also presents problems in the form of impeding holistic judgements, and the overall moderation process adds to the assessors' workload, limits their assessment options, and that the increase in reliability appears limited.

The various literature cited above appear to suggest that while the concept of moderation has been part of the assessment sphere for some time, it has only recently received more focus from the assessment bodies, Australia in particular (Bloxham, Hughes, & Adie, 2015; Adie, Lloyd, and Beutel, 2013), and thereby the increase in the number of studies revolving around assessment moderation. What many researchers and studies appear to agree on is the role of criteria in assessment and by association, moderation. Studies have found that the use of criteria is a highly valued part of moderation (Van der Schaaf, Baartman, and Prins, 2012; Wyatt-Smith, Klenowski, and Gunn, 2010; Adie, Klenowski, and Wyatt-Smith, 2012) perhaps due to the role of moderation as a means of ensuring reliability and enforcing accountability.

2.9 Studies Involving Moderation

Bloxham, Hughes, and Adie (2015) relates a collaboration in moderation research between UK and Australia with the aim of identifying the concept of moderation and its approaches, and to examine its advantages and potential within the assessment realm. They examined the purpose of several moderation practices: peer scrutiny of assessment design, second marking, consensus/social moderation discussion, consideration of grade distributions, and external peer review. In examining these practices, the authors also

considered what roles the four discourses of equity, justification, accountability and community building have within the practices. They found that most of the practices examined revolved around the purposes of accountability and justification; equity, while an important purpose of moderation, was less apparent through their examination of evidence – and may even be at odds with practices revolving around consistency (grade distribution, averaging marks in second marking). The purpose of community building was found to aid in establishing a shared knowledge of standards, and aid assessors in understanding the reasons for moderation. The authors acknowledge the complexity of moderation and note that many universities appear to have a limited understanding of moderation practices but the authors agree that the purpose of moderation is quality assurance.

Orr (2007) differentiates between the positivist and the poststructuralist approaches towards assessment research and attempts to demonstrate the poststructuralist nature of moderation conversations. She examined 10 hours of observation of moderation practices on the “assessment of art and design studio-based undergraduate work” (p. 649) and found that lecturers reached mark agreement through an exchange of rhetorical moves. In addition, she found that when there was a disagreement, the lecturers had a tendency to split the difference, or would move the mark closer to one score – which she cites as Dracup’s (1997) “the agreement process”. She indicated the presence of power relations in the moderation process, pointing out that the subjects would defer to experience, and that the more experienced, or the subject leaders were less prone to budge on the marks they had given. She further found that marks were lowered or raised as an attempt to influence the students’ future efforts and performance. Orr’s limited data appears to place moderation well within the poststructuralist perspective of assessment.

Bloxham (2009) highlights several assumptions regarding assessment in higher education institutions, these include 1) the ability to accurately and reliably mark students' work, 2) internal moderation ensures fair marking standards, 3) external moderation ensures consistency of standards across institutions, and 4) students' final award is a consistent reflection of achievement both within and across institutions. She disputes assumption 1, pointing out the inherently interpretative nature of assessment, likening the assessors to 'connoisseurs', and argued that despite the existence of criteria and marking standards assessment is a negotiation of a social construct influenced by tacit knowledge. The author acknowledges the possibility of moderation as a measure for ensuring fairness and its role as a mitigation tool, but points again to the inherent interpretative nature of assessment in the interpretation of assessment criteria, and highlighted the role of power relation thus disputing assumption 2. Bloxham questions the effectiveness of assumption 3, pointing to the increasingly complex and modularized nature of programs in institutions of higher education and argued that the external examiner may be limited by their own experience. The author debunks assumption 4, citing reports showing the inconsistencies of degree classifications within and across institutions. Bloxham lamented the burden of work these apparently ineffective moderation practices place on assessors and attempts to provide alternatives to this work heavy practice. She considered the feasibility of emulating the assessment practices of the school sector but admits that the rigid marking schemes, extensive assessment training, and controlled examination conditions would not be suitable in the university setting. She then explored the possibility of acknowledging and accepting the subjectivity in assessments and discarding the notion of a 'correct' mark, thus limiting the need for moderation to inconsistent or borderline performances, new assessment formats, and to help develop mutual understanding of criteria. She also suggested the possibility of a module level approach where patterns of differences in assessment among different groups of assessors are

identified. She further examined the possibility of involving students in the assessment process by encouraging them to assess, therefore making the assessment process transparent and simultaneously making it part of the student's learning process.

Adie, Lloyd, and Beutel (2013) reviewed, through semi-structured interviews, the moderation practices of 25 academic teaching staff in an Australian University's Faculty of Education and identified four distinct "discourses of moderation" (p. 972). The moderation discourse of equity related to fairness and consistency in assessment, the discourse of justification was seen in the assessor's confidence in their marking and their ability to justify said marking decisions, the discourse of community building is a collaborative process where the assessors developed shared understandings of standards, criteria, tasks etc., and the discourse of accountability referred to the normative distribution of marks and grades. While the discourses may be distinct, the authors stressed that moderation is an overlapping, interconnected process which is highly complex and that beliefs and practices of moderation differed by individual.

Van der Schaaf, Baartman, and Prins (2012) in examination of moderation in practice, found that teachers are not necessarily objective during moderation, but are social and respond to one another by complementing each other, and agreeing with or supporting what others have said. They indicated that while the use of criteria does regulate reliability in assessment and moderation, the assessments are sometimes based on personal opinion as it was found that the teachers do not necessarily examine the evidence thoroughly or critically but were found to reach agreement on a consensus with little difficulty, suggesting perhaps that teachers can be and are swayed by one another in the moderation setting.

The above studies focus on several aspects of moderation and its role in assessment, and while they touch on its role (Bloxham, 2009), the various types of moderation (Bloxham, Hughes, & Adie, 2015), the different theoretical approaches to moderation (Orr, 2007), the motivations behind moderation and how assessors perceive the role of moderation (Adie, Lloyd, & Beutel, 2013), the role of criteria in assessment and by extension moderation (Van der Schaaf, Baartman, & Prins, 2012; Wyatt-Smith, Klenowski, and Gunn; 2010), and even the interpretation of criteria and the understanding of the terminology and qualifiers in said criteria (Wyatt-Smith, Klenowski, and Gunn; 2010; Adie, Klenowski, & Wyatt-Smith, 2012), it appears however that there are few studies on the discourse within the moderation process but for the brief study found in Orr (2007) to illustrate how moderation in practice differs from the positivist approach aimed at ensuring reliability and accountability in moderation.

2.10 Application of Argumentation to Moderation

This study will seek to examine the discourse that occurs during the paired moderation process while the assessors attempt to come to an agreement on the marks they had individually awarded to students in an assessed performance. To do so, the study will make use of Van Eemeren and Grootendorst's (2004) pragma-dialectic framework, particularly the Ideal System for Resolution of Disputes to examine the discussion between the paired moderators as they seek to reach a consensus. The ideal system appears to be a highly suitable tool for analyzing and evaluating moderation discourse given the goal of moderation, for the desired result is that the moderating assessors agree on one score, and that the ideal system is a guideline on examining and evaluating discussions where the outcome is one single conclusion. And given the speech act basis underlying pragma-dialectics, the use of this framework will allow for the analysis of

another dimension in the form of moves made by the moderating assessors as well as their underlying speech acts.

Pragma-dialectics also appears to be highly suited to the examination of moderation discourse for it is a discourse based analytical tool and according to Jorgensen and Philips (2002), discourse analysis allows for the examination of language use in various situations, including moderation discourse. Jorgensen and Philips further point out that in discourse based analysis, the analyst is limited to what is present within the analyzed discourse and that the analyst is not to determine truth or intentions. This is highly parallel to pragma-dialectical analysis where the objective is to examine what is said or expressed, and from there analyze and evaluate the discourse, the arguments put forth, and how the argument is concluded and how solid the arguments and the resulting conclusion. At no point does the pragma-dialectical analyst ascribe intentions to the discourse, however the framework does to some extent try to determine truth, as seen in the focus on logic and fallacies.

Since this study intends to examine the discourse generated during the moderation process and how a consensus is reached, and given the nature of moderation, which is not debate based nor particularly rhetorical or argumentative in nature, it will not focus on the logic and fallacies aspect of pragma-dialectics. Instead, the study will make use of an adapted version of the quality of argumentation classification found in Van der Schaaf, Baartman, and Prins (2012), which is based on Clark and Sampson (2007) and Erduran, Simon, and Osborne (2004). This study adapted and operationalized the categories (refer Appendix B or Table 4.27) provided by Van der Schaaf, Baartman, and Prins (2012) to better suit the data as the versions in Clark and Sampson (2007) and Erduran, Simon, and Osborne (2004) were more suited to scientific discourse based on evidential argumentation.

Moderation, as has been seen in the discussion above, has several positive aspects, among these the enforcement of standards through the use of criteria, and its role as a means of ensuring reliability. However, moderation has its drawbacks, it appears that power relations play a role, where the balance skews towards the senior assessor, as Orr (2007) and Reiman et al. (2010) show. It appears, another worrying trend is that of moderators agreeing on scores with sparse justifications or evidence (Orr and Bloxham, 2013), and this appears to contradict Sadler's (2013) argument for making use of evidence found within the candidate's work. Yet another point of concern in the practice of moderation is the use of grades as reference markers, as Barrie, Hughes, and Crisp (2014) point out, Yorke (2008) too laments the unfortunate consequences of norm-referencing.

2.11 Conclusion

Having explained the various theories and concepts in this research, and some principles on how the analysis of the data will be carried now, the next chapter will discuss the methodology that was applied in this study.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter will describe the methodology in which this study was carried out. It will include details about the theoretical framework, the research design, the participants, data collection procedures, and data analysis. The objective of the study is to examine the talk assessors engage in during the process of coming to a consensus on the scores awarded to four students in a group oral interaction, and how they justify this judgement. The study draws strongly on the pragma-dialectic argumentation theory, which is a discourse based framework, making it a discourse analysis study and thus mainly qualitative in nature.

3.2 Theoretical Framework

Van Eemeren and Grootendorst's (2004) pragma-dialectic framework is applied in this study and it is a framework which allows for the study of both the disputation process and the reasoning behind the disputation. However, the study will be based mainly on the Ideal System for Resolution of Disputes which provides a guideline for the analysis and evaluation of a dispute. Additionally, the quality of argumentation classification found in Van der Schaaf, Baartman, and Prins (2012) will also be applied to evaluate the arguments put forth by the participating assessors. Pragma-dialectics was not able to fully draw out the evaluation of the arguments in the data due to the framework's focus on logic on fallacies. As the study does not seek to examine the fallacies or argumentative loopholes the moderators may have committed, nor does the data allow for a sufficient study of logic and fallacies, it was deemed appropriate that the use of the quality of argumentation

classification be employed here. A comprehensive write-up of both the pragma-dialectic framework and the quality of argumentation classification has been given in Chapter 2.

3.3 Research Design

The study makes use of data from the UMRG research project titled “Employers’ Assessment of Language Competence in Group Oral Interaction” (project number: RP014B-13SBS). The participants had individually watched and assessed a videotaped group oral interaction between four students, and were later paired for moderation based on the differences in the scores they had given. In the moderation, the paired participants were briefed on the need to come to an agreement on the scores for each of the four students in the video. As the data consists of the talk between the assessors, the approach taken to analyze the data is discursive. Hence this study is qualitative in nature as it seeks to make sense of the spoken discourse in the disputation process, the linguistic and/or pragmatic aspect in the dispute, and the evaluation of the assessors’ respective arguments.

3.4 Data

Four paired moderation sessions that were audiotaped and transcribed were analyzed. Two of these, Pair 1 and Pair 2 were analyzed in detail, while the other two, Pair 3 and Pair 4 were used to support, consolidate, or disprove the findings from Pairs 1 and 2. Refer Table 3.1 for details of the paired moderations used in the study.

Table 3.1 Moderation Pairs and Details

Moderation session	Length of moderation	Moderators
Pair 1	41 Minutes	A1, B1
Pair 2	26 Minutes	A2, B2
Pair 3	24 Minutes	A3, B3
Pair 4	21 Minutes	A4, B4

3.5 Participants

The participants were selected from members of the teaching staff from both the English Language Department and the Language Unit of the Faculty of Languages and Linguistics, University of Malaya. A survey was distributed to all the teaching staff of the two departments/units. The selection criterion was the respondent's experience in assessing group oral interaction. Of the 52 respondents, 34 met this criterion. When approached, 14 of this number volunteered to participate in this study. The 14 were further broken down into two groups: novice and experienced assessors, where the former comprise of assessors with less than 20 years of experience in teaching and assessment of 10 or less courses with a group oral interaction component, and the latter consisting of assessors with 20 or more years of experience in teaching and assessment of more than 10 courses involving a group oral component. Table 3.2 provides some background information on the assessors.

All participants signed consent forms indicating their willingness to participate in the study and agreed to allow the collected data to be used for research purposes. They were also aware that they were being audio recorded and that the recording would constitute as data. Permission to use the data was obtained from the principal investigator of the

project. For privacy and anonymity, the participants have been assigned arbitrary designations.

Table 3.2 Participant Profiles for Moderating Assessors

Assessor	Assessor Role	Experience	
		As a Teacher or Assessor	Assessing group oral interaction
A1	Experienced	20 years	22 courses/semesters
A2	Experienced	28 years	12 courses/semesters
A3	Experienced	30 years	15 courses/semesters
A4	Experienced	28 years	11 courses/semesters
B1	Novice	10 years	8 courses/semesters
B2	Novice	9 years	3 courses/semesters
B3	Novice	12 years	4 courses/semesters
B4	Novice	3 years	2 courses/semesters

3.6 Data Collection

The participants were interviewed individually and during this interview, they were asked to watch and assess a videotaped group oral interaction involving four students from the course GTEE1109 titled “Speaking Skills in English”. The participants assessed and scored the students in the video using the rating scale and marksheet from the course (Refer to Appendix C). The participants were then paired for assessment on two criteria:

- i) the differing scores they had awarded the four students in the video,
- ii) their experience in assessing group oral interaction (their identities as novice or experienced assessors).

Before they began the moderation sessions, the participants were briefed on the need to come to an agreement on the scores they had awarded each of the four students. Additionally, they were asked to discuss the reason(s) for having awarded the respective

scores for each student, including that for the scores for the three subsections. They were also given the option of reviewing the video at any time should they wish to do so. The moderation was audio recorded and transcribed using an adaptation of the Conversation Analysis Transcription Notation (Refer Appendix D).

3.6.1 Video Taped Group Oral Interaction

The data for this study was obtained in various ways. A group oral interaction among four students from the course GTEE1109 was recorded. For a close reflection of their probable performance during their actual assessment in the group oral component, the video was taken a few days before their actual assessment date to ensure that they had received the full benefit of instruction within the group oral component. The project's researchers devised a test situation that closely reflected those of the course which was then vetted by the course coordinator (Refer to Appendix E). The students participating in the video were then numbered N1, N2, N3, and N4 respectively from left to right for both ease of identification and anonymity. Consent was obtained from the students for the use of the videotaped interaction for research purposes. Running length of the video was 22 minutes.

3.7 Data Analysis

Given the objective and research questions of the study, two paired moderations have been selected for detailed analysis. These two pairs (Pair 1 and Pair 2) were selected because the participating assessors appear to have more active exchanges in addition to having presented more extensive justification for the scores they had awarded the students. Within the two selected pairs, only a selected segment of the moderation will be analyzed based on the research questions above. The selected segment consists of the

beginning of the moderation, up until each pair has reached their initial consensus for the total score of the first student they have chosen to discuss. This boundary was identified and enforced as it is felt that the respective selected segment will adequately address the research questions, and provide sufficient data for analysis.

To answer RQ 1, the distinct discussions/arguments in the selected segments are first identified, then the discussions are then compared against the Ideal System for Resolution of Disputes. Here, the stages of discussion, namely the confrontation stage, opening stage, argumentation stage and concluding stage, are identified for each distinct discussion found within the selected segment. The similarities and differences between the moderation discourse and the ideal system are then identified and discussed.

Since much of the selected segment in both pairs appears to comprise the argumentation stages of each distinct discussion as identified earlier:

- 1) the argumentation stages seem to encompass most of the justifications or defenses presented, and
- 2) it also appears that the argumentation stages are where the assessors convince each other to come to a consensus;

hence RQ 2 and RQ 3 focus on the argumentation stages present in the selected segments.

To answer RQ 2, each argumentation stage identified in RQ 1 is analyzed to identify what moves the assessors make (i.e. advance argument, seek clarification, express acknowledgement, advance explanation, express agreement etc.), and what speech acts each move comprises.

Following this, to answer RQ 3, the arguments identified in RQ 2 are then analyzed through a classification for Quality of Arguments as adapted from Van der Schaaf, Baartman and Prins (2012), using Baartman (2007) as a classification guideline. This is to determine the ‘quality’ of the arguments the assessors put forth either to defend or justify the score they had awarded, or argue against the score their opponent had awarded.

Only after analysis for Pair 1 and Pair 2 were complete, and patterns and generalities identified and drawn, were Pair 3 and Pair 4 analyzed. The patterns found in Pairs 1 and 2 were searched for in Pairs 3 and 4, the resulting findings were then discussed. Additionally, general and superficial patterns and findings from the entire moderation session were then identified and discussed briefly.

3.8 Unit of Analysis

This is a qualitative study based primarily on the analysis of the moderation discourse found within two paired moderation sessions. The analysis of this study, as mentioned in the previous section, is guided by three research questions which are in turn based on Van Eemeren et al.’s (1993) pragma-dialectic Ideal System for Resolution of Disputes, and the Quality of Argumentation classification found in Van der Schaaf, Baartman and Prins (2012). As the three research questions examine different aspects of the moderation discourse, it would then follow that the unit of analysis would vary for all three questions.

Under the first research question, the data is compared against the Ideal System for Resolution of Disputes (refer Table 2.1 or Appendix A) for both similarities and differences. The unit of analysis utilized in the analysis for this question is segments of moderation data which approximates one complete discussion of a difference in opinions

(scores) between the paired assessors from the identification of said difference of opinion, to where both assessors are in agreement of the score they will award the student. This complete discussion will subsequently be referred to as a distinct discussion. A distinct discussion refers to a stretch of discourse revolving around one particular topic with a clear purpose, in this case, the airing of a difference in opinion and the ensuing exchange leading up to and including an agreement regarding this difference. For instance, the difference in opinion is identified when one assessor reveals that she has given 9 marks and the other reveals that she has given 7 marks. These different scores constitute the difference in opinion. These scores or marks which they reveal they had given are also their respective standpoints for the corresponding discussions, i.e. the former assessor's standpoint is 9 marks, and the latter assessor's standpoint is 7 marks. The assessors would then discuss their reasons or justifications for having awarded their respective marks, and at the end, reach an agreement on the mark they both find to be acceptable for the student. As this question requires the moderation data to be compared to the ideal system, the segment of a complete discussion of difference in opinions, is then analyzed in order to identify the confrontation stage, opening stage, argumentation stage, and concluding stage found in that distinct discussion. In order to find these stages in the data, key moves characteristic to each discussion stage is identified to determine the presence of the discussion stages in a distinct discussion. For instance, any expression of standpoint (scores) would signify the presence of a confrontation stage, any challenge to defend said scores would indicate an opening stage, all advancements of arguments and counter arguments would denote an argumentation stage, and the establishment of results or the determination of an outcome (scores) would bespeak a concluding stage.

The second research question examines the moves and speech acts that the assessors use in the argumentation stage. As was explained in section 3.7, the argumentation stage is

chosen as the analysis focus of the second and third research questions as it comprises a large part of the data, it encompasses most of the justifications and defenses presented, and it appears to be where the assessors attempt to convince each other to adopt their stance. Under this research question, only the argumentation stages identified under the first research question will be analyzed for moves and speech acts. In order to identify these moves and speech acts, each discrete utterance is evaluated as a move in terms of the role it plays in the context of the discussion. These moves may include *advance argumentation* (or *put forth argument*), *seek clarification*, *express acknowledgement*, *advance explanation*, *express agreement* etc. For instance, in Excerpt 1.1, B1 and A1 justify the marks they had each had given Student Number 4 (N4) for Language Control, Vocabulary and Structures (LC). The moves identified here comprises one move of *putting forth argument* (assertive) by B1 in Lines 115-116, one move of *expressing agreement* (commissive) and two moves of *putting forth argument* (assertive) by A1 in Lines 117-118. B1's utterance in Line 119 is considered a backchannel and is thus not counted as a move or speech act in this analysis. Then in Lines 120-121, A1 continues with one move of *put forth argument* (assertive), to which B1 adds in Line 112, one move of *put forth argument* (assertive) which A1 responds to in Line 123 by *expressing agreement* (commissive). Altogether, A1 makes 3 moves of *put forth argument* and 2 moves of *expressing agreement*, i.e. 3 assertives and 2 commissives while B1 makes 2 moves of *put forth argument*, resulting in a count of 2 assertives.

Excerpt 1.1 Pair 1 Transcript (Lines 115-123)

115	B1	Because I thought that her language, she actually made some:: obvious
116		errors (),
117	A1	Yeah, I've got- marked down that she's got (.) formulaic expressions,
118		ah:: fossilized errors and all, showing up::,
119	B1	°Mm::,°
120	A1	And:: yeah. There's that typical: Malaysian English ah:: you know
121		verb 'be' missing,

122	B1	A lot of fillers::,
123	A1	That's right, yeah, and::

The third research question then examines the arguments identified in the second research question, and evaluates them using the operationalized Quality of Argumentation classification adapted from Van der Schaaf, Baartman and Prins (2012) as seen in Table 4.27 in Section 4.4 (also in Appendix B). To illustrate, each move of *put forth argument* identified from Excerpt 1.1 above is then evaluated using the Quality of Argumentation. Following the operational definition, both B1's moves of *put forth argument* are classified as Level 3 arguments because B1 uses evidence from the video in her argument, but the evidence presented is in the form of general observation. A1's three moves of *put forth argument* on the other hand, are classified as Level 4 arguments, this is because A1 makes more specific observations from the video, where B1 observes that there were language errors, A1 identifies the kinds of language errors, thus citing more specific evidence from the video.

3.9 Reliability of the Study

As the section on Unit of Analysis demonstrates, there is a degree of complexity to the analysis in this study, which demands proof of the credibility of the results. This proof is all the more important in this study which seeks, to some extent, to evaluate how assessors justify their judgement of a group oral interaction assessment. To this end, a pilot study had been conducted on one paired moderation session to evaluate the feasibility of the study and the analytical tools used. The team involved in the research project RP014B-13SBS co-rated the analysis for the pilot study based on the unit of analysis spelled out above along with the two analytical tools of the Ideal System for Resolution of Disputes (see Appendix A) and the Quality of Argumentation classification (see Appendix B). The

project team comprised three lecturers from the Faculty of Languages and Linguistics (one now retired) and one project consultant (formerly a lecturer with the Faculty of Languages and Linguistics). Their areas include discourse analysis, language learning, and teaching and assessment. Using the respective units of analysis for the three research questions, it was found that the team achieved a high level of inter-rater agreement for the three research questions, though the percentage of agreement was never measured. Ducasse and Brown (2009, p. 432) cites Storch (2001) as stating that “in discourse studies, the level of agreement is often in the vicinity of 80% of the data coded”. In the rare instances where agreement was not reached, the researcher and the co-raters discussed the instances of non-agreement, identified the reasons for not reaching agreement, then further refined the analytical procedure where possible.

3.10 Conclusion

The sections above have described the methodology, starting with theoretical framework, followed by research design, data, participants, data collection procedure, data analysis, unit of analysis, and the reliability of the study. The next chapter will provide the analysis of the data.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 Introduction

The objective of this study is to examine how assessors in moderation sessions reach a consensus in the scores they had awarded, and in what way they justify how they had assessed the students. To achieve this objective, the data will be analyzed based on the three research questions which guide this study. As mentioned in Chapter 3, two paired moderations will be analyzed in detail, and analysis will focus on the selected segment spanning the beginning of the moderation to agreement of total score for the first student chosen for discussion. The following sections present the findings obtained from the analysis.

4.2 Analysis on the Ideal System for Resolution of Disputes

To answer the first research question: “In what way does the moderation discourse between assessors compare to Van Eemeren et al.’s (1993) pragma-dialectical Ideal System for Resolution of Disputes”, the distinct discussions within the respective selected segments of both paired moderations are identified. A distinct discussion refers to a stretch of discourse revolving around one particular topic with a clear purpose, in this case, the airing of a difference of opinion and the ensuing exchange leading up to and including an agreement regarding this difference. Each distinct discussion is then compared against the Ideal System for Resolution of Disputes where the four stages of discussions are identified should they be present. The stages of discussion are identified based on moves or purposes particular to the discussion stage, for instance the expression

of standpoint(s) for confrontation stage, discussion of discussion rules or common ground for opening stage, the advancement of argument(s) for the argumentation stage, and the establishment and agreement on results/conclusion for concluding stage. The similarities and differences between the data and the ideal system is then identified and discussed.

4.2.1 Pair 1 (Assessors A1 and B1)

In the selected segment of lines 1-289 (for full transcript, refer to Appendix F), Pair 1 appeared to discuss briefly who they respectively found to be the best student in the group oral interaction, before moving on to a discussion of scores. They appeared to have reached an implicit agreement to discuss N4's scores and proceeded to discuss the scores they had given N4 in TF (Task Fulfillment), LC (Language Control), and CA (Communicative Ability) respectively. Four distinct discussions were discerned from the selected segment, these being:

- 1) The discussion on best performer in group,
- 2) The discussion on N4's TF score,
- 3) The discussion on N4's LC score,
- 4) The discussion on N4's CA score.

The stages of discussion were then identified for each of these four distinct discussions, where it was found that the stages of discussion occur multiple times and out of order in all four of Pair 1's distinct discussions, as demonstrated in Tables 4.1 to 4.4. Due to space constraints, the segments of transcripts are not provided here, but will be identified for each of the respective discussions and can be referred to in Appendix F.

4.2.1.1 Discussion on Best Performer

The moderation discourse of Pair 1 begins with A1 who starts the discussion by asking which of the four students B1 thought the best performer in the group was. B1 identifies N2 and N3 (confrontation stage), then justified her choice of N3 (argumentation stage). B1 then acknowledges N4's performance upon encountering doubt (argumentation stage). A1 then reveals that she thought N4 had been the best (confrontation stage) and endorsed B1's arguments for N4 (opening stage) and provided further arguments for N4 (argumentation stage) before reiterating her standpoint. A1 then suggested that they look at the scores they had each given the students (concluding stage). This analysis has been tabulated and put into a table form as presented in Table 4.1.

Table 4.1: Discussion stages and their order in the discussion on best performer

Sequence of discussion	Discussion stage
Express standpoint	Confrontation stage
Advance argumentation	Argumentation stage
Express standpoint	Confrontation stage
Establish common ground	Opening stage
Advance argumentation	Argumentation stage
Suggest means of resolution (accepted)	Concluding stage

The identification of the stages is based on the argumentation model proposed by Van Eemeren et al (1993). It is observed that there are instances when the analysis reflects the model adopted for this study, however there are also instances where there are deviations from this model. More analysis as follows will explain these instances of similarities and differences in detail.

4.2.1.2 Discussion on N4's TF Score

The talk between assessors A1 and B1 for N4 has evidence of how the pair resolves their differences. For a start, Pair 1 implicitly agree to discuss N4's scores and started with the TF score. It was B1 who begins by revealing that she had given 9 marks (confrontation stage) and then provided minimal defense of this (argumentation stage). A1 then revealed that she had given 7 marks (confrontation stage) and proceeded to provide argument against the score B1 had given (argumentation stage). To this, B1 made minimal attempts to defend her score but appears unsuccessful (argumentation stage). A1 then suggested they moderate their marks by accepting middle ground, B1 accepts this course of action. This brought about the agreement on a score of 8 for N4's TF score (concluding stage). Table 4.2 tabulates the discussion stages.

Table 4.2: Discussion stages and their order in the discussion on N4's TF score

Sequence of discussion	Discussion stage
Express standpoint	Confrontation stage
Advance argumentation	Argumentation stage
Express standpoint	Confrontation stage
Advance argumentation	Argumentation stage
Suggest means of resolution (accepted) Establish result	Concluding stage

4.2.1.3 Discussion on N4's LC Score

Moving on from the TF score, A1 and B1 now reveal the scores they had given in the LC section, 9 and 7 marks respectively (confrontation stage). As the unfolding discourse shows, A1 attempted to point out that they had both given marks in the same score banding (opening stage). B1 proceeded to justify the score she had given and A1 supports

the arguments and provided further arguments of her own (argumentation stage). B1 then notes that the marks they had given were within the same score banding (opening stage). A1 then states that 8 marks was also in the same score banding (concluding stage). Pair 1 did not finalize on 8 marks, instead the pair revisited the discussion of their shared score banding and noted the descriptors in the scoring guide for the score band (opening stage). Finally, B1 offers the suggestion of meeting halfway again and both agree on 8 marks for N4's LC score (concluding stage). Table 4.3 shows the stages for this discussion.

Table 4.3: Discussion stages and their order in the discussion on N4's LC score

Sequence of discussion	Discussion stage
Establish differences	Confrontation stage
Establish shared common ground (attempted)	Opening stage
Justify scores given	Argumentation stage
Establish shared common ground	Opening stage
Suggest means of resolution (not acted on)	Concluding stage
Established shared common ground discuss scoring guide descriptors	Opening stage
Suggest means of resolution (accepted) Establish results	Concluding stage

4.2.1.4 Discussion on N4's CA Score

For the CA score, A1 states that she had given 7 marks while B1 reveals she had given 5 marks (confrontation stage). A1 advanced arguments to defend the 7 she had given, some of which had B1 agreement (argumentation stage). Then B1 attempts to defend her score by highlighting what she felt was a weakness on N4's performance, and to this A1 provided counterarguments in the form of explanations incorporating scenarios, some of which did not involve observations of N4's performance (argumentation stage). This

received a rejection from B1, who maintains her argument on N4's performance (argumentation stage), followed by a suggestion that they reach a compromise of 6 marks, to which A1 responds with a counteroffer of 6.5 marks. This was rejected by B1 (concluding stage).

This caused A1 to question B1's refusal to accept and A1 suggested reviewing the group oral interaction (opening stage). B1 defended her refusal of the 6.5 mark solution by citing her earlier argument, and then A1 argued that there were other criteria in the CA section (argumentation stage). B1 then revises the score she had given to 6 marks, A1 rejects 6 marks as a solution and upholds her 7 marks, forcing a compromise, resulting in N4 receiving 6.5 marks for CA score (concluding stage). A tabulation of the discussion stages follows in Table 4.4.

Table 4.4: Discussion stages and their order in the discussion on N4's CA score

Sequence of discussion	Discussion stage
Establish differences	Confrontation stage
Advance argumentation	Argumentation stage
Suggest means of resolution (rejected)	Concluding stage
Issue challenge	Opening stage
Establish possible common ground	
Advance argumentation	Argumentation stage
Suggest means of resolution (partly accepted)	Concluding stage
Establish results	

4.2.1.5 Further Analysis of Pair 1 Findings

Having identified the stages of discussion in each distinct discussion, further analysis showed that some of the identified discussion stages do not conform fully to the respective

discussion stages in the ideal system. This is seen in the absence of certain moves listed within the ideal system (Refer Appendix A). One example is the opening stage of the discussion on best performer (see Excerpt 1.2, Lines 43-44, in italics) and is reproduced below for easy reference. There is no explicit challenge to defend the standpoint issued, as well as no discussion or agreement regarding discussion rules, instead, A1 merely expressed agreement with the argument B1 had advanced for N4, and in doing so, established common ground.

Excerpt 1.2 Pair 1 Transcript (Lines 43-49)

43	A1	And then I found 4:::, actually stands out. <i>Precisely for the same</i>
44		<i>reasons that you said, she is able to lead the discussion,</i> and what I
45		found, uhm::: most impressive about her was that, she was able to bring
46		the discussion back to focus. When- when one of them, had gone off
47		tangent, she actually was the one who brought it back into focus, so I
48		would think, in terms of her speaking ability, and also this being group
49		interaction,

It is found that all of the concluding stages found in the analysis of Pair 1’s discourse do not to conform to the concluding stage in the ideal system. In all four distinct discussions, A1 and B1 never once upheld their standpoints, nor did they accept their opponent’s standpoint. Instead, results were established through compromise, or settlement. It is also noted that in the discussion on N4’s TF score (Excerpt 1.3), A1 suggests a means to resolve their difference in scores where she cites that in moderation practice it is conventional that the marks are averaged.

Similar to the discussion on N4’s TF score, the discussions on N4’s LC and CA scores show the same deviation from the format of the concluding stage proposed in the ideal system. Both these discussions indicate that this pair withdrew their standpoints without accepting or adopting their counterpart’s standpoints. Instead, they both appear to agree

on a new score somewhere between the scores they had initially awarded. This is evident in the discourse shown in Excerpt 1.3.

Excerpt 1.3 Pair 1 Transcript (Lines 96-107)

96	A1	So if we are moderating, and there's like 2 marks of difference,
97	B1	Mmhmm,
98	A1	I'm willing to go up by 1,
99	B1	8
100	A1	Yeah,
101	B1	8 marks. Ok, that's ok. I'll go down by 1.
102	A1	[So that means you go down
103	B1	Ok,
104	A1	Ah 1. So Number 4, Student Number 4, the first part, we agree on 8,
105		yeah?
106	B1	[Number
107	B1	Yeah, 8 marks.

Additionally, in the concluding stage for the discussion on best performer, no result was established, but only a proposal of a means of resolving the difference which was subsequently accepted as neither A1 nor B1 was willing to relinquish their respective standpoints even after an exchange of arguments. This particular concluding stage could be classified as an opening stage as well, in the sense that A1 and B1 are considering the possibility of resolution through a discussion of scores. This could be accepted as an agreement of discussion rules. Hence, Excerpt 1.4 could be classified both as a concluding stage as well as an opening stage despite not fulfilling the parameters of either stage. It fits the parameters of an opening stage in that they agree to begin discussion, and the 'discussion rules' being a focus on the scores they had given. However, no challenge was issued to defend standpoints, and no discussion in which to establish existing common ground was put forward. Similarly, it only partially fits the parameters of a concluding stage in that it marks the end of the discussion on best performer in group in favor of moving on to another discussion. There is no concrete result that A1 or B1 agree on whom

the best performer was and neither party withdraws their standpoint. In fact, the difference was not resolved at all, instead it was deferred by the suggestion of a discussion of scores and this could be due to the fact that this is part of the opening stage of the entire moderation session between A1 and B1.

Excerpt 1.4 Pair 1 Transcript (Lines 56-58)

56	A1	And maybe we should look at the marks we've given,
57	B1	Ok,
58	A1	And then see how far away (.) these are.

4.2.1.6 Summary of Pair 1 Discussion Stages

Table 4.5 provides a tabulated summary of the discussion stages Pair 1 had engaged in for each of the four discussion stages analyzed. The table also indicates the number of times Pair 1 had visited the stages in each discussion and has made indication whether or not Pair 1's discourse had conformed to the discussion stages found in the Ideal System for Resolution of Disputes. For instance, it can be seen that under Pair 1's discussion of the best performer in the group, there were two Confrontation and Argumentation stages respectively as opposed to only 1 Opening stage and 1 Concluding stage. This indicates that Pair 1 had not followed the order of discussion stages suggested in the Ideal System for Resolution of Disputes in this particular discussion. It can also be observed from the table that in this discussion of best performer in the group, the confrontation and argumentation stages were highly similar and thus fully conformed (FC) with the description of these discussion stages as suggested in the ideal system; while the opening and concluding stages only partially conform (PC) with those suggested in the ideal

system, hence indicating differences between these stages found in the data and that in the ideal system, as had been demonstrated earlier.

Table 4.5: Pair 1 discussion stages

Discussion/stages	Confrontation			Opening			Argumentation			Concluding		
	FC	PC	#	FC	PC	#	FC	PC	#	FC	PC	#
Best performer	/	-	2	-	/	1	/	-	2	-	/	1
N4's TF	/	-	2	-	-	-	/	-	2	-	/	1
N4's LC	/	-	1	-	/	3	/	-	1	-	/	2
N4's CA	/	-	1	-	/	1	/	-	2	-	/	2

(key: FC – Fully conform, PC – Partially conform, # - number occurrences)

4.2.2 Pair 2 (Assessors A2 and B2)

In the selected segment of Lines 1-234 (refer to Appendix G for full transcript), Pair 2 appear to exchange views on the order of students, from best to worst, in terms of their performance in the group oral interaction, then proceed to discuss the scores of N2 in terms of TF, LC, and CA. Four distinct discussions were found in the selected segment of Pair 2's moderation, these are:

- 1) The discussion on best performer in group,
- 2) The discussion on N2's TF score,
- 3) The discussion on N2's LC score,
- 4) The discussion on N2's CA score.

The stages of discussion in each distinct discussion was then identified, and it was found to be similar to the discursual pattern used by Pair 1. It is noted also that Pair 2's four distinct discussions did not comply with the suggested order of discussion stages proposed in the ideal system, and this is shown on Tables 4.6 to 4.9. Not only were there

multiple occurrences of certain discussion stages, two of the discussions did not have the supposedly pre-requisite argumentation stage, and one of these two discussions did not have a concluding stage. As mentioned above, the segments of transcripts are not provided due to space constraints, but can be found in Appendix G. Segments relevant to each discussion will be indicated.

4.2.2.1 Discussion on Best Performer

The moderation discourse of Pair 2 is initiated by A2 who starts the discussion by asking B2 which student she thought had performed best in the group. B2 reveals that she found N2 to be the best performer (confrontation stage), which A2 agreed with (opening stage). A2 then asked B2's to rank the remaining students in terms of performance in descending order which B2 reveals to be N4, N1, and N3 respectively (confrontation stage). A2 states her agreement with B2's ranking of the students' performances, then suggested they looking the students' 'banding' (opening stage). Table 4.6 presents the tabulation of the discussion stages.

Table 4.6: Discussion stages and their order in the discussion on best performer

Sequence of discussion	Discussion stage
Express standpoint	Confrontation stage
Establish shared common ground Suggest means of starting discussion	Opening stage
Further expression of standpoints	Confrontation stage
Establish shared common ground Suggest of starting discussion	Opening stage

4.2.2.2 Discussion on N2's TF Score

In relation to the analysis of Pair 1's discourse on the TF scores for N4, the following is the analysis of Pair 2's discussion on another student (N2) who they agreed was the best speaker in the group oral task. It begins with A2 prompting B2 who then reveals that she had given 7.5 marks to the candidate (confrontation stage). A2 then reveals that she had given 7 marks (confrontation stage). This is followed by A2 stating that they can average the marks they had given later (concluding stage). A tabulation of the discussion stages can be found in Table 4.7.

Table 4.7: Discussion stages and their order in the discussion on N2's TF score

Sequence of discussion	Discussion stage
Establish differences	Confrontation stage
Suggest means of resolution	Concluding stage

4.2.2.3 Discussion on N2's LC Score

Next, upon prompting, B2 reveals that she had given 10.5 marks for N2's LC score, A2 reveals that she had given 8 marks (confrontation stage). B2 then questions A2's mark (opening stage). This is then followed by A2's arguments, where she defends the score she had given (argumentation stage). She also referred to the score descriptors for the score band for the marks as well as that which B2 had given (opening stage). A2 then revised her score to 9 marks, prompting B2 to revise her score to 10. As a result of this exchange of talk, this pair then agree to average their revised marks, resulting in N2 receiving 9.5 marks for LC score (concluding stage). The discussion stages are tabulated in Table 4.8.

Table 4.8: Discussion stages and their order in the discussion on N2's LC score

Sequence of discussion	Discussion stage
Establish differences	Confrontation stage
Challenge to defend standpoint	Opening stage
Advance arguments	Argumentation stage
Referred score descriptor (discussion rules)	Opening stage
Withdraw standpoint Suggest means of resolution (accepted) Establish result	Concluding stage

4.2.2.4 Discussion on N2's CA Score

For the Communicative Ability section, A2 revealed that she had given 5 marks while B2 allotted 8.5 marks (confrontation stage). To this B2 questioned A2 on the mark she had given (opening stage), and B2 took it one step further by providing argumentation against A2's mark, citing the criteria for CA in the process (argumentation stage). A2 acknowledges B2's arguments and explained that she had overlooked some of the criteria, then advanced arguments in defense of her own score before arguing against B2's score (argumentation stage). Then A2 conceded that the 5 marks she had given was very low while B2 acknowledged that 8.5 marks was too high, A2 then suggested that they moderate their marks (concluding stage). A2 once again defends her score by stating she had overlooked part of the criteria, then argued that the criteria did not make allowances for one aspect of N2's performance (argumentation stage). This prompted a quick reference to and a brief discussion on the criteria (opening stage). A2 again voices doubts about upholding the 5 marks she had given (concluding stage). This then prompted B2 to remind A2 of one of the latter's earlier arguments, which A2 then expands on, garnering brief counterarguments from B2 (argumentation stage). Eventually Pair 2 agreed on one

of A2's arguments (argumentation stage). At the end, A2 revised her score to 6 marks, to which B2 responds with 6.5, upon prompting, B2 reveals that 6.5 is the final average, thus N2 receiving 6.5 marks for CA score (concluding stage). Table 4.9 presents a tabulation of the discussion stages.

Table 4.9: Discussion stages and their order in the discussion on N2's CA score

Sequence of discussion	Discussion stage
Establish differences	Confrontation stage
Challenge to defend standpoint	Opening stage
Advance arguments	Argumentation stage
Retract standpoint Suggest means of resolution (not accepted)	Concluding stage
Advance arguments	Argumentation stage
Refer to scoring guide (discussion rules)	Opening stage
Offer to retract standpoint	Concluding stage
Advance arguments	Argumentation stage
Suggest means of resolution (accepted) Establish results	Concluding stage

4.2.2.5 Further Analysis of Pair 2 Findings

Similar to Pair 1, not all moves suggested for each stage of discussion could be found within Pair 2's moderation. While a few of the identified stages do display all the suggested moves, others conform only minimally to the stages proposed in the ideal system. This is especially so for the opening stages of the discussions on N2's LC score and N2's CA score, the multiple opening stages collectively display all the moves suggested in the ideal system, these moves are interspersed with argumentation stages and even a concluding stage. It appears that a challenge to defend a standpoint is issued in the initial opening stage, then the pair moves on to the argumentation stage (and

beyond) before revisiting the opening stage to establish or perhaps review discussion rules.

Excerpt 2.1 is an excerpt of talk between the two assessors on who each found to be the best performer in the group oral interaction. The segment of talk appear to simultaneously be a confrontation stage and an opening stage as it contains moves salient to both discussion stages.

Excerpt 2.1 Pair 2 Transcript (Lines 1-32)

1	A2	So we're going to discuss the marks (.) that er:: we're going to discuss
2		as a team, ah,
3	B2	Mmhmm,
4	A2	For finalization. So:: according to the four speakers, uhm wh-who do
5		you think is the best?
6	B2	For me I still find that- ah for me I find that the second speaker, (.) is
7		better,
8	A2	[Is the best
9	B2	[Is the best.
10	A2	among the whole group lah.
11	B2	Yes.
12	A2	Ya. So I also gave the highest marks there,
13	B2	Mmhmm,
14	A2	So uhm (.) you want to discuss it (.) ah:: one by one? As in:: ok, we- we
15		better maybe we put it in order of bet-
16	B2	The best.
17	A2	The best and then the second, the third, the fourth,
18	B2	Mmhmm,
19	A2	Then maybe we (.) look into the small marks, ah?
20	B2	Mmhmm.
21	A2	Ok, so:: Speaker 2 is the best,
22	B2	Mmhmm,
23	A2	Then followed by?
24	B2	Fourth. The fourth speaker.
25	A2	Ya, same also, my same:: ah opinion there, and then?
26	B2	Mmhmm, and then followed by the:: I have the First Speaker,
27	A2	As the better one than the::
28	B2	The Third One,

29	A2	Ya. Same.
30	B2	Mmhmm.
31 32	A2	Same with mine. Ok, so::: now we go into their::: sort of banding lah, hor,

It appears that the confrontation stage and the opening stage occur simultaneously in the discussion on best performer in group as can be seen in Excerpt 2.1. In Lines 6-11, B2 reveals upon prompting that N2 is the best performer, then in Line 12, A2 reveals that that too is her standpoint, thereby expressing her standpoint and establishing the common ground between them. This is repeated when B2 revealed that she found N4 to be the second best performer in the group, and again when the rest of the order in which they have ranked the students is revealed. While the instances in which B2 expresses her standpoints may perhaps be considered stand alone confrontation stages on their own, that they are so intertwined with A2's revelations that those too are her standpoints make it such that the entire exchange on standpoints can be classified simultaneously as both confrontation stage and opening stage. Trying to delineate the two stages would perhaps result in an extremely disjointed and confounding exchange.

Like Pair 1, all of Pair 2's concluding stage do not conform to the concluding stage found in the ideal system. All three of Pair 2's discussions involving numerical scores see A2 and B2 coming to a compromise on the scores they had initially given the students. In the discussion on N2's TF score, it appears that A2 and B2 would average their respective scores for a result both agree on. Similarly, in the discussions on N2's LC and CA scores, they again appear to compromise on a final score somewhere between their initial scores. The final concluding stage for the discussion on N2's CA score appears particularly interesting as shown in Excerpt 2.2. Here, like all other discussions involving scores in this analysis, A2 revised her score, thereby withdrawing her standpoint and putting forth a new one. However, unlike the other two discussions on scores between Pair 2 and the

three discussions on score between Pair 1, B2 did not stand by her score or standpoint before both parties agree to a compromise. Nor did she revise her score before the two moderating assessors decide to average their scores. Instead, it appears that B2 has skipped that step and has put forth a score which the pair accepts as the final result of their discussion.

Excerpt 2.2 Pair 2 Transcript (Lines 202-216)

202	A2	Ya. Ya. Ok, I agree. Ya. Too low::: Because generally now that I see,
203		I'm more on the stricter side, and you're the more lenient, and::: as a
204		group, as a whole, I feel that ah, we- we- is- is- we can discuss this lah.
205		Tsk,
206	A2	[So::: I would probably bring up to::: six, you know,
207	B2	[ok, we'll see
208	A2	Or six point five.
209	B2	Six point five lah.
210	A2	So and you?
211	B2	Six point five.
212	A2	You also bring down to six point five?
213	B2	No lah, I mean we take the average (.) for the total,
214	A2	So that means,
215	B2	Six point five lah.
216	A2	Ok.

Additionally, due to the nature of Pair 2's discussion on best performer, there was no attempt to establish results or a conclusion at all, as can be seen in Excerpt 2.1. This could perhaps be because there is no difference to resolve. As mentioned above, both A2 and B2 establish their standpoints, and simultaneously agree on these standpoints, thus establishing common ground. Hence, it appears there is no concluding stage as there is no need to discuss or establish results.

4.2.2.6 Summary of Pair 2 Discussion Stages

Table 4.10 provides a tabulated summary of the discussion stages Pair 2 had engaged in for each of the four discussion stages analyzed. The table also indicates the number of times Pair 2 had visited the stages in each discussion and has made indication whether or not Pair 2's discourse had conformed to the discussion stages suggested in the Ideal System for Resolution of Disputes. It can be seen that all the confrontation and argumentation stages conform fully while none of the opening and concluding stages do.

Table 4.10: Pair 2 discussion stages

Discussion/stages	Confrontation			Opening			Argumentation			Concluding		
	FC	PC	#	FC	PC	#	FC	PC	#	FC	PC	#
Best performer	/	-	2	-	/	2	-	-	-	-	-	-
N2's TF	/	-	1	-	-	-	-	-	-	-	/	1
N2's LC	/	-	1	-	/	2	/	-	1	-	/	1
N2's CA	/	-	1	-	/	3	/	-	3	-	/	3

(key: FC – Fully conform, PC – Partially conform, # - number occurrences)

4.2.3 Summary of Findings for Analysis Based on the Ideal System

The detailed analysis of the discourse between the two pairs of moderators has shown how the data may not be completely similar to the Ideal System for Resolution of Disputes proposed by Van Eemeren et al. (1993) when compared. It appears from Tables 4.5 and 4.10 that the each of Pair 1 and Pair 2's four respective discussions do not fully conform to the argumentation model. It can be observed that all the confrontation stages of the discussions by both pairs conform fully to the format suggested within the ideal system, as do the argumentation stages, however, none of the opening stages or concluding stages conform to that which is suggested within the ideal system. Deviations from the suggested

concluding stage has been described extensively in the analysis above. Deviations from the opening stages appear in the form of the absence of certain salient moves attributed to the opening stage. While it can be seen in the analysis of each of the discussions by both pairs of moderating assessors that they do perform certain moves salient to the opening stage, every one of the opening stages found in the analysis display only one of these signature moves, two at best. The moderating assessors appear to skip some of these moves or perform them implicitly. While it would be an interesting avenue of research to pursue how exactly the opening stages found within the analysis do not conform to that suggested in the ideal system, it does not provide enough for analysis in this study. Instead, the focus of further analysis would be on the argumentation stages found in the discussions of both pairs as these do not appear to deviate from the format suggested in the ideal system. Furthermore, there appears to be more exchanges in the argumentation stages, providing more opportunity for further analysis. More importantly, the argumentation stages appear to be where the final result or conclusion is determined as the arguments put forth by the participants in the argumentation stages function to convince and persuade their respective opponents to accept their standpoint.

4.3 Moves and Speech Acts Used in Argumentation Stage to Justify Scores Given

This section of the discussion will focus on what was put forth as RQ2, and for easy reference it is repeated here: “in the argumentation stage(s), what moves and speech acts do the assessors use in justifying the score given”. Focus is put on the argumentation stage because this is where the participants of a critical discussion often persuade one another of the acceptability or otherwise of a particular standpoint. The analysis begins with identifying the moves that the assessors employ in the argumentation stages (examples of these moves are *putting forth argument*, *expressing agreement*, *seeking clarification* and

so on). Moves and speech acts were the chosen focus of this research question given that the study makes use of the pragma-dialectic framework which is in turn anchored by speech act theory. After the moves are identified, the speech acts involved in the making of these moves are then categorized. This then leads to a phase where a comparison is made among the assessors (of both pairs) to determine if any pattern of use should surface.

4.3.1 Pair 1 (Assessors A1 and B1)

In the analysis from the four distinct discussions of Pair 1's (A1 and B1) talk of lines 1-289 (refer to Appendix F for full transcript), it appears that they engage in the discussion stage of argumentation a total of 7 times. These consist of:

- 1) Discussion on best performer: 2 argumentation stages
- 2) Discussion on N4's TF score: 2 argumentation stages
- 3) Discussion on N4's LC score: 1 argumentation stage
- 4) Discussion on N4's CA score: 2 argumentation stages

The 7 argumentation stages identified within these 4 discussions suggest a tendency on the part of Pair 1 to put forth arguments, move on to another stage of discussion before revisiting the argumentation stage and establishing a result both parties agree on. With the exception of the discussion on N4's Language Control (LC) score, it appears that the second argumentation stage in each discussion leads to the final conclusion. Subsequently an examination of the moves exchanged within the argumentation stages was carried out. These moves are broken down to the various component moves such as arguments or sub-arguments, and the speech acts underlying these component moves are identified. For instance, in B1's arguments justifying having chosen N3 as the best performer, each trait

mentioned counts as 1 move and is therefore 1 speech act (assertive). Due to space constraints, only the tabulated summaries of Pair 1's discussion on best performer in group will be shown in this chapter. All subsequent tabulated summaries for Pair 1 under the moves and speech act research question can be found in the Appendix H.

4.3.1.1 Discussion on Best Performer

The argumentation stages in this discussion span the lines 12-37 and lines 39-55 (refer to Appendix F for transcript). In the first argumentation stage (lines 12-37), there are 2 main arguments:

- B1's justification for choosing N3
- B1's acknowledgement of N4's merits

In the second argumentation stage (lines 39-55), there is only 1 main argument, and that is A1's justification for choosing N4. Each of these three main arguments consists of component moves where A1 or B1 puts forth arguments or sub-arguments to support or substantiate the main argument. However, the actual analysis of the arguments will only be done in RQ3, as this RQ focusses on moves and speech acts. Other than arguments, moves such as the expression of agreement, acknowledgement, questions of clarification, or the acceptance of an argument will also be included.

In the moderation discourse where the two assessors are deliberating on who is the best performer in stage 1 of the argumentation phase (lines 12-37), B1 puts forth three arguments justifying her choice of N3. A1 then tries to pin down B1's standpoint and ends up expressing doubt before asking after B1's standpoint once more. B1 repeats standpoint, then acknowledges N4's merits (5 arguments). A1 expresses agreement on

N4's merits. In this stage, it appears A1 uses, 2 commissives, 1 directive, and 1 usage declarative but no assertives while B1 uses 10 assertives and 2 commissives. Table 4.11 provides a summary of the moves and speech acts used by Pair 1 in this argumentation stage.

Table 4.11: Summary of moves by speech act in argumentation stage 1 of discussion on best performer (Lines 12-37)

Speech acts used	Purpose of speech act	
	A1	B1
Assertive A1: - B1: 10	-	Justify choice of N3 (3) Acknowledge N4's merits (5) Repeat/Reconfirm standpoint (2)
Commissive A1: 2 B1: 2	Expressing doubt (1) Agree on N4's merits (1)	Repeat/reconfirm standpoint (2)
Directive A1: 1 B1: -	Ask for B1's standpoint (1)	-
Usage declarative A1: 1 B1: -	Attempt establish B1's standpoint (1)	-

Moving on to stage 2 of the argumentation phase (lines 39-55), A1 begins by stating her opinion that N2 and N4 were of similar levels before further stating that N4 stood out. She then proceeded to give 3 arguments before reiterating that N4 stood out. In brief, A1 uses 7 assertives while B1 makes no discernible moves. A summary of the moves and speech acts in this argumentation stage is tabulated in Table 4.12.

Table 4.12: Summary of moves by speech act in argumentation stage 2 of discussion on best performer (Lines 39-55)

Speech acts used	Purpose of speech act	
	A1	B1
Assertive A1: 7 B1: -	State standpoint – N2, N4 (2) Statement – N4 stand out (2) Put forth arguments – N4 (3)	-
Commissive A1: - B1: -	-	-
Directive A1: - B1: -	-	-
Usage declarative A1: - B1: -	-	-

4.3.1.2 Discussion on N4's TF Score

From the discussion on the best performer, the analysis will now move to the TF (Task Fulfillment) score of N4's. The argumentation stages within this discussion occur at line 63 and at lines 67-95 (refer to transcript in Appendix F). The former consists of a very brief argument from B1 defending her standpoint while the latter is an argument from A1 against B1's standpoint, which B1 attempts to refute. In the stage 1 of the argumentation, B1 had expressed her standpoint (TF=9) which appeared to have been met with doubt, leading her to defend it minimally (1 assertive). There were no other moves by either A1 or B1. Next moving on to stage 2 of the argumentation stage, A1, after having expressed her own standpoint (TF=7), proceeded to put forth arguments against B1's standpoint (2 assertives). It is apparent that B1 does not readily accept A1's arguments, instead using a total of 4 directives to clarify the topic of discussion and to question or confirm the criteria

under TF. A1 then clarifies the topic and the TF criteria before citing the scoring guide as part of her argument against B1's standpoint. B1 appears to acknowledge A1's arguments but does not appear to accept them. In this stage, A1 used 7 assertives and 8 commissives while B1 used 2 assertives, 4 commissives, and 4 directives. Table 4.13 in Appendix H provides a summary of the speech acts and moves used within this argumentation stage (stage 2).

4.3.1.3 Discussion on N4's LC Score

To move on, the next section in the assessment criteria, i.e. Language Control (LC) score is examined for N4 whose scores both A1 and B1 had agreed to discuss first. There is only one argumentation stage in this and it occurs at lines 115-123. Here, A1 appears to agree with and corroborate the arguments B1 puts forth in defending her standpoint. In short, A1 used 3 assertives and 2 commissives while B1 uses only 2 assertives. A summary of the moves and speech acts for this argumentation stage can be found in Table 4.14 in Appendix H.

4.3.1.4 Discussion on N4's CA Score

Finally for N4's Communicative ability (CA) score, it is observed that there are two argumentation stages and these are evident in lines 163-230 and lines 265-278. In the former, A1 defends her standpoint (CA=7) with arguments, justifications and explanations and then counters B1's attempts to argue against a higher CA score. In the second argumentation stage of this discussion, A1 puts forth arguments in an attempt to convince B1 to relinquish her standpoint of CA=5.

For argumentation stage 1 (Lines 163-230, Appendix F), A1 puts forth arguments on communicative qualities to defend her standpoint (5 assertives) and further mentioned the groups' performance in terms of fluency and pronunciation (2 assertives). B1 expresses agreement, then questioned A1 regarding N4's hesitation. A1 clarifies her stand regarding N4's hesitation through explanations and justifications, leading B1 to question if fillers and hesitations were synonymous. A1 provided an explanation on her view regarding fillers and hesitations then related this with her interpretation of N4's performance. B1 then argued that N4 had a lot of fillers and language errors. A1 acknowledges this, further stating that she (A1) had marked N4 down for language errors. Here, A1 makes use of 26 assertives, 4 commissives and 2 directives. B1 on the other hand used 6 assertives, 10 commissives and 2 directives. A tabulated summary of the moves and speech acts in this argumentation stage can be found in Table 4.15 in Appendix H.

Moving on to the second argumentation stage in this discussion (Lines 265-278, Appendix F), B1 defends her refusal to accept the compromise of CA=6.5 with the argument that 'I feel that she wasn't very fluent' in Line 265 (expressive). A1 counters this, pointing out that CA had 2 other components besides fluency and that N4's performance deserved higher marks for those. Here, A1 used 5 assertives and 1 commissive while B1 did not make use of any assertives, instead using 2 commissives and 1 expressive. Table 4.16 in Appendix H provides a summary of the moves and speech acts used in this argumentation stage.

4.3.1.5 Comparison of Speech Acts Used by A1 and B1

Table 4.17 shows the speech acts that A1 and B1 use in the argumentation stages found within the four discussions under analysis. Here, it can be seen that A1 uses more than

double the number of assertives B1 does. The number of commissive both A1 and B1 use appear similar while B1 uses twice the number of directives A1 does.

Table 4.17: Summary of speech acts by A1 and B1

SPEECH ACTS IN THE 4 DISCUSSIONS	A1	B1
TOTAL ASSERTIVES	48	21
Discussion on best performer	7	10
Argumentation stage 1	-	10
Argumentation stage 2	7	-
Discussion on N4's TF	7	3
Argumentation stage 1	-	1
Argumentation stage 2	7	2
Discussion on N4's LC	3	2
Discussion on N4's CA	31	6
Argumentation stage 1	26	6
Argumentation stage 2	5	-
TOTAL COMMISSIVES	17	18
Discussion on best performer	2	2
Argumentation stage 1	2	2
Argumentation stage 2	-	-
Discussion on N4's TF	8	4
Argumentation stage 1	-	-
Argumentation stage 2	8	4
Discussion on N4's LC	2	-
Discussion on N4's CA	5	12
Argumentation stage 1	4	10
Argumentation stage 2	1	2

Table 4.17, continued

TOTAL DIRECTIVES	3	6
Discussion on best performer	1	-
Argumentation stage 1	1	-
Argumentation stage 2	-	-
Discussion on N4's TF	-	4
Argumentation stage 1	-	-
Argumentation stage 2	-	4
Discussion on N4's LC	-	-
Discussion on N4's CA	2	2
Argumentation stage 1	2	2
Argumentation stage 2	-	-
TOTAL EXPRESSIVES/USAGE DECLARATIVES	1 usage declarative	1 expressive
Discussion on best performer	1 usage declarative	-
Argumentation stage 1	1 usage declarative	-
Argumentation stage 2	-	-
Discussion on N4's TF	-	-
Argumentation stage 1	-	-
Argumentation stage 2	-	-
Discussion on N4's LC	-	-
Discussion on N4's CA	-	1 Expressive
Argumentation stage 1	-	-
Argumentation stage 2	-	1 Expressive

4.3.2 Pair 2 (Assessors A2 and B2)

In the four distinct discussions within the selected segment of Lines 1-234 (refer Appendix G for transcript), Pair 2 appear to have engaged in a total of 4 argumentation stages in only two of these discussions, these being:

- 1) Discussion on N2's LC score: 1 argumentation stage
- 2) Discussion on N2's CA score: 3 argumentation stages

While the discussion on N2's LC score has only one argumentation stage, the following discussion on N2's CA score has 3, suggesting that the pair may have had difficulty reaching agreement, as seen by the sequence of discussion stages under RQ1 findings, where A2 and B2 appear to have to refer to the discussion rules and even attempted to establish results only to revisit their arguments before an agreement was achieved. This then calls for an examination of the moves exchanged within the argumentation stages. These moves are broken down to the various component moves such as arguments or sub-arguments and so on, after which the speech acts underlying these component moves are identified. Similar to B1's arguments justifying having chosen N3 as the best performer in Pair 1's moderation, each trait mentioned counts as 1 move and is therefore 1 speech act (assertive).

The way the analysis is done mirrors what has been carried out for Pair 1. Due to space constraints, detailed analysis of Pair 2's speech acts as used in the argumentation stages will be summarized below and the table can be found in Appendix I.

To begin this section on N2's Language Control (LC) score, it is noted that the sole argumentation stage in this discussion spans from lines 52-67 (refer Appendix G for transcript) which consisted mostly of A2 accepting the challenge to defend her standpoint (LC=8) and advancing arguments to that end, garnering agreement from B2. Here, A2 had used 5 assertives and 1 commissive while B2 used only a single commissive (refer to Table 4.18 in Appendix I for a summary of the moves and speech acts in this argumentation stage).

Moving on to the discussion on N2's Communicative Ability (CA) score, it was found that there are three argumentation stages and they are as follows:

- 1) Argumentation stage 1: Lines 101-153
- 2) Argumentation stage 2: Lines 161-176
- 3) Argumentation stage 3: Lines 178-201

For Stage 1 argumentation in Pair 2's discourse on N2's CA score (Lines 101-153, Appendix G), A2 begins by responding to B2's challenge indicating that their score difference is down to her (A2) being more strict. B2 does not accept this, and cites the scoring guide as part of her argument against A2's standpoint. A2 appears to accept B2's argument, then explains that she (A2) had referred only to the scoring guide and not the CA appendix of the rating scale (refer to Appendix C) before putting forth arguments, citing that N2's fluency did not reflect the level expected at the score of CA=8.5 and reasons this out. Here, A2 uses 20 assertives and 5 commissives while B2 uses 13 assertives and 4 commissives. Refer to Table 4.19 in Appendix I for a summary of the moves and speech acts used in this argumentation stage.

For the next argumentation stage (Stage 2), see Lines 161-176 of Appendix G, A2 once again indicates the differences between the scoring guide and the CA appendix of the rating scale (refer to Appendix C), then argued that N2 had over led and how this had affected the group. The pair then briefly referred to the CA criteria and made observations on the group in view of the criteria being discussed. In this argumentation stage, A2 used 10 assertives and 2 commissives while B2 used 1 assertive and 4 commissives. Refer to Table 4.20 in Appendix I for a summary of the moves and speech acts used in this argumentation stage.

In argumentation Stage 3 (Lines 178-201, Appendix G), B2 appears to remind A2 of her argument regarding N2's over leading and its effects, to which A2 puts forth further argument based on her observation of N2's actions and supported these through N2's failure to prompt others in the group. B2 countered mildly, garnering an admission from A2 that she can't remember before A2 furthers her argument, stating N2's failure to prompt the quiet ones, which B2 agrees and supports. A2 further argued that N2 held a dialogue with N4 which prompted a counterargument from B2 that N4 had prompted others in the group. A2 acknowledges this and redirects the discussion back to N2 and tried to rationalize N2's actions before arguing that N2 had kept the discussion going. B2 then argued against A2's standpoint which ultimately led to the final concluding stage of this discussion. In the final argumentation stage within Pair 2's discussion on N2's CA score, A2 used 7 assertives and 4 commissives while B2 used 6 assertives and 7 commissives. A tabulated summary of the moves and speech acts used in this argumentation stage can be found in Table 4.21 in Appendix I.

Table 4.22 shows the speech acts that A2 and B2 use in the argumentation stages found within the four discussions under analysis. Here, it can be seen that A2 uses more than double the number of assertives B2 does. It appears that B2 uses a third more commissives than A2 does.

Table 4.22: Summary of speech acts used by A2 and B2

Speech acts in the 4 discussions	A2	B2
TOTAL ASSERTIVES	42	20
Discussion on N2's LC score	5	-
Discussion on N2's CA score	37	20
Argumentation stage 1	20	13
Argumentation stage 2	10	1
Argumentation stage 3	7	6
TOTAL COMMISSIVES	12	16
Discussion on N2's LC score	1	1
Discussion on N2's CA score	11	15
Argumentation stage 1	5	4
Argumentation stage 2	2	4
Argumentation stage 3	4	7
TOTAL DIRECTIVES	-	-
TOTAL EXPRESSIVES/USAGE DECLARATIVES	-	-

4.3.3 Types of Moves and Speech Acts

Table 4.23: Comparison of the speech acts used by Pair 1 and Pair 2

Speech act	Pair 1		Pair 2	
	A1	B1	A2	B2
Assertive	48	21	42	20
Commissive	17	18	12	16
Directive	3	6	-	-
Expressive	-	1	-	-
Usage declarative	1	-	-	-

From Table 4.23, it can be observed that while the 'A' assessors use more than double the number of assertives that the 'B' assessors do, both 'A' and 'B' assessors use approximately similar numbers of commissives with 'B' assessors having uttered a token number more than the 'A' assessors. It appears that B1 has used double the number of

directives A1 has. It can be observed that A1 has used one usage declarative in her attempt to confirm B1's standpoint in the discussion on best performer in group, and B1 uses an expressive in defending her decision not to agree to the solution A1 suggests to resolve their differences in the discussion on N4's CA score. However, usage declaratives and expressives, and even directives appear not to be categories of speech acts that Pairs 1 and 2 use often in the respective selected segments used for analysis.

4.3.3.1 Assertives

The analysis so far has shown that the various moves have been categorized with the speech act types and that the 'A' assessors (A1 and A2) put forth more arguments and make more explanations and observations than the 'B' assessors (B1 and B2), thus making use of more assertives. While it is apparent that the 'A' assessors put forth more arguments than the 'B' assessors, the breakdown in Table 4.24 shows that A1 puts forth nearly three times the number of arguments that B1 does at 32 and 11, whereas A2 only puts forth half again the number of arguments that B2 does at 17 and 11 respectively.

It appears from the moves used by Pair 1 that A1 may be the one in charge of the discussion in that A1 puts forth more arguments, A1 makes explanations while B1 does not, A1 makes more statements, and A1 is the one who establishes the topic(s) under discussion which B1 does not do at all. While B1 does voice more observations than A1, the move of voicing observations appears not to carry as much weight as making explanations, and establishing topic. Therefore, it may be surmised here that A1 was in charge of the discussion in Pair 1 from the greater number of assertives uttered and arguments put forth in addition to the moves of making explanations and establishing topic. It may also be surmised from the greater number of assertives uttered, the greater

number of arguments, statements and observations made by A2 compared to B2, in addition to the moves of making explanations and establishing topic that A2 seems to be in charge of the discussion within Pair 2. See Table 4.24 for the categorization of the moves used by Pairs 1 and 2 involving assertives.

Table 4.24: Categorization of assertives by moves made by Pair 1 and Pair 2

Moves	Pair 1		Pair 2	
	A1	B1	A2	B2
Put forth argument	32	11	17	11
Make explanation	4	-	10	-
Make statement	4	2	8	2
Make observation	1	6	5	2
Repeat standpoint	2	2	-	-
Establish topic	4	-	1	-
Prompting	1	-	-	3
Counterclaim	-	-	-	2
Correction	-	-	1	-
TOTAL	48	21	42	20

4.3.3.2 Commissives

In terms of commissives, it can be observed in Table 4.25 that A1 expresses agreement with more frequency than B1 does whereas in Pair 2, B2 appears more active than A2 in expressing agreement. In terms of acknowledging what has been said, it appears that the 'B' assessors do so more than the 'A' assessors. In fact, A1 does not at any point within the argumentation stages found within the selected segment acknowledge what has been said, instead she appears to use firmer forms of responses such as the expression of agreement or confirmation. From the moves in the category of the speech act category of commissives used by the assessors in Pair 1, it appears that A1 is more firm in her interactions as observed by the moves she uses while B1 appears a little less committed

in the moves she utilizes under the speech act category of commissives. In Pair 2, B2 appears to express agreement and acknowledgement with more frequency than A2, B2 also appears to make more decisive responses such as accepting or not accepting an argument, prompting, and discussing discussion rules. A2 makes such moves in the form of redirecting their discussion and correcting what has been said, but A2 also makes conceding moves such as expressing admission or admitting shortcoming, and conceding arguments. Hence it appears that A2 has taken a more conciliatory approach in terms of the moves made under the speech act category of commissives.

Table 4.25: Categorization of Commissives by Moves made by Pair 1 and Pair 2

Moves	Pair 1		Pair 2	
	A1	B1	A2	B2
Express agreement or disagreement	10	4	3	5
Express acknowledgement	-	7	2	6
Respond to prompt or question	1	1	1	-
Express confirmation	3	1	-	-
Accept or not accept argument	-	1	-	1
Prompt	1	-	-	3
Express doubt	1	1	-	-
Express admission	-	1	2	-
Correction	-	-	1	-
Concede argument or point	-	-	2	-
Redirect discussion	-	-	1	-
Discuss discussion rules	-	-	-	1
Repeat standpoint	1	2	-	-
TOTAL	17	18	12	16

4.3.3.3 Directives

As can be seen in the summary and the breakdown in Table 4.26, B1 uses twice the number of directives A1 does, however, these appeared in the form of seeing clarification

and questions. At no point did B1 issue a direct challenge. Pair 2, it appears, does not use any directives in the selected segments used for analysis.

Table 4.26: Categorization of Directives by Moves made by Pair 1 and Pair 2

Moves	Pair 1		Pair 2	
	A1	B1	A2	B2
Ask question	1	4		
Seek clarification	1	2		
Pose tag question	1	-		
TOTAL	3	6		

Having established in Table 4.23 that the ‘A’ assessors use more than double the number of assertives than the ‘B’ assessors do, most of this discrepancy stemming from the number of arguments that the ‘A’ assessors put forth and the explanations given as well as the statements made, it appears that an evaluation of the various arguments and justifications put forth by the assessors of Pair 1 and Pair 2 would then be prudent. This then leads to the analysis for the third research question of the study.

4.4 Structure and Presentation of Justifications

This section will present the findings from the analysis on how the justifications given by the assessors are structured and presented, which is the focus of the third research question in this study. Each claim made in support of the assessors’ respective standpoint is first identified, after which the arguments and justifications put forth to support this claim is identified and its relation to the claim and other arguments determined. Having established how the arguments made in support of the respective claims are structured, the arguments, or support, is then analyzed using the Quality of Argument classification adapted from Van der Schaaf, Baartman, and Prins (2012) as can be seen in Table 4.27.

As presented in the table, the levels of classification have been defined and operationalized, and the arguments or supports are then classified following this operational definition. In doing so, the ‘quality’ of the arguments the assessors offered is then evaluated to examine if, perhaps, the ‘quality’ of arguments influence the outcome of the discussion in any way.

While Level 1 and Level 5 arguments are relatively easy to recognize, the former in its absence or irrelevance to the claim and the latter in its depth and complexity (subject to relevance). There are some challenges when it comes to identifying if an argument is of Level 2 or 3, in addition to the separation of the Level 3 and Level 4 arguments. Hence an approach employed to differentiate between these argument levels is found in Baartman’s (2007) classification of arguments (refer to Appendix B), and is therefore adapted for this analysis. It is suited in this context as references to the rating scale in support of a claim would then be classified as Level 2, because the rating scale functions as a guideline for assessment as well as discussion rules for the moderation. Evidence encompasses observations and comments from the assessors as they viewed the video-taped performance of the group oral interaction. To differentiate between Levels 3 and 4 arguments, both of which hinge on evidence from the video as support, it was determined that Level 3 arguments would then consist of simple observations or comments of the assessed performance while Level 4 arguments comprise more specific or detailed observations from the assessed performance, or possibly simple observations which are then elaborated upon.

Table 4.27: Quality of argumentation classification

Level 1	No support <i>Nothing or irrelevant reason(s) is given to support the standpoint, score, or claim</i>
Level 2	Using explanation as support <i>A brief explanation is used to support the standpoint, score or claim</i>
Level 3	Using evidence as support <i>Evidence from the assessed performance is used to support the standpoint, score or claim</i>
Level 4	Using evidence as support <i>Further or more precise evidence from the assessed performance is used to support the standpoint, score or claim</i>
Level 5	Coordinating multiple pieces of evidence or multiple connections between ideas in the evidence <i>Multiple pieces of evidence from the assessed performance is used to demonstrate how a conclusion is drawn to support the standpoint, score or claim</i>

(adapted from Van der Schaaf, Baartman, & Prins, 2012, p.853)

4.4.1 Pair 1 (Assessors A1 and B1)

As was found in the analysis involving the Ideal System for Resolution of Disputes, Pair 1 had engaged in the argumentation stage for all the four discussions under analysis.

These being:

- 1) The discussion on best performer in group,
- 2) The discussion on N4's TF score,
- 3) The discussion on N4's LC score, and
- 4) The discussion on N4's CA score.

For the analysis given in the following subsections, this is the key used: C stands for Claim, S for Support and Q for Qualifier.

4.4.1.1 Discussion on Best Performer

To begin with this section, the discussion is taken from Pair 1 where assessor B1 puts forth two sets of arguments, the first in defending N3 and therefore defending her own standpoint, the second in putting forth N4's merits. A1 on the other hand puts forth only one set of arguments, which is to defend N4. The Excerpt 1.5 (Lines 12-23) show how the discourse unfolds, where B1 presents her arguments in support of N3.

Excerpt 1.5 Pair 1 Transcript (Lines 12-23)

12	B1	Because for- for- actually, initially Number 3, ah she wasn't- she
13		didn't:::
14	B1	[stand out
15	A1	[say much, ah?
16	B1	Ah, she didn't stand out. she didn't say much in the beginning as well.
17		But towards the end, I think she participated more and uhm,
18	A1	Right.
19	B1	And:: after that, you can see that her:: eye contact,
20	A1	Mmhmm?
21	B1	She really participated in- in the discussion. And ah:: perhaps less
22		grammatical errors, I mean, she does have, but (.) perhaps less
23		grammatical errors, that's what I noticed,

As shown in the excerpt, the argument offered by B1 is analyzed as:

N3 is the best speaker in the group: this is a claim labelled C1.1

- Didn't stand out/say much initially: this is a qualifier (Q1)
- Participated more towards end: this is a support (S1)
- Eye contact present (S2) - really participated (S2.1)
- Had less grammatical errors (S3)

The arguments are evaluated below:

Q1, S1 : Level 3
(Simple observations which were not elaborated upon)

S2, S2.1 : Level 4
(Evidence from video, eye contact was linked to participation. Not Level 5 because it was only 1 point)

S3 : Level 3
(Simple observation which is not elaborated on, nor were examples given)

The structure used in defending N3 is: at Level 3 – 2 occurrences, at Level 4 – 1 occurrence.

Later on B1 offers arguments in support of N4's performance. N4 is not B1's standpoint, but she (B1) only acknowledges N4 due to A1's doubts about N3. This is based on Lines 29-36 (refer to Appendix F for transcript). The analysis is given below:

B1 acknowledges N4's merits (C1.2), and according to the transcript, she mentions:

- Has leadership skills (S1)
- Very attentive (S2) – had right body language (S2.1)
- Involved everyone (S3)
- Led discussion (S4)

The arguments are evaluated as follows:

S1 : Level 2, bordering Level 3
(Explanation inferred from what was seen in video, no evidence or example given, the precise skills for instance)

S2, S2.1 : Level 4
(Evidence from video, linked attentiveness and body language. Not Level 5 because it was only 1 point)

S3 : Level 3
(Evidence from video, no elaboration such as how she involve everyone)

S4 : Level 3, bordering 4
(Observation made from evidence from video. No elaboration or example of leadership given)

The structure used in acknowledging N4's merits is: at Level 2 – 1 occurrence, at Level 3 – 2 occurrences, and at Level 4 – 1 occurrence.

Following this, A1 then offered arguments in support of N4 in Lines 43-54 (refer to Appendix F for transcript), the analysis follows:

A1 states that N4 stands out (C1.3), and according to the transcripts, she cites:

- Same reasons you said (S1) - Led discussion (S1.1)
- Bring discussion back to focus (S2)
 - o When someone had gone off tangent, she brought it back to focus (S2.1)
- Excellent skills in listening to others' opinion (S3)
 - o Speaking ability in group interaction (S3.1)
 - Brought discussion back to focus (S3.1.1)

The arguments are evaluated below:

S1, S1.1 : Level 2
(Expresses agreement with B1's argument. No fresh evidence added)

S2, S2.1 : Level 4
(Evidence from video, explained how N4 brought discussion back to focus)

S3, S3.1, S3.1.1: Level 5
(Linked listening with speaking in group interaction context, segued to this from the point of bringing disc. back to focus with the linker 'so')

The structure in A1's arguments in support of N4's performance is: at Level 2 – 1 occurrence, at Level 4 – 1 occurrence, and at Level 5 – 1 occurrence.

From the analysis provided above, it is noted that there are a total of 3 claims and 10 arguments offered by Pair 1 in support or defense of N3 and N4. This is tabulated in Table 4.28.

4.28: Summary of the Level of Arguments by Pair 1 (BP)

Level of argument	A1	B1
Level 2	1	1
Level 3	-	4
Level 4	1	2
Level 5	1	-

Having considered the justifications on the best performer, the analysis will move now to the various sections on the scoring guide, and to begin the argument for N4's Task Fulfilment (TF) score.

4.4.1.2 Discussion on N4's TF Score

There are four sets of arguments, the first being B1's argument in defense of her standpoint, then A1's argument against B1's standpoint, after which is B1's argument against A1's standpoint, and finally A1's argument that N4 deserves only the second scoring band for her performance.

B1 started the discussion by defending the score of 9 she had given N4 for TF (Lines 61-63, Appendix F), the analysis is below:

B1 suggested that N4 deserves 9 for TF (C1.4), stating:

- I think she did a good job (S1)

The argument is evaluated as follows:

S1 : Level 2
(Explanation and expression of personal opinion, no evidence from video)

The structure in defending TF=9 is one occurrence of Level 2 argument.

Following this, A1 argued against B1's standpoint of TF=9, as seen in Lines 67-78 (refer Appendix F). The analysis follows:

A1 argued that N4 does not deserve 9 for TF (C1.5), reasoning that:

- TF A to C grades, 9 is A+ (S1)
- N4 not A material, N4 is at A- level (S2)

The arguments are evaluated below:

S1 : Level 2
(Explanation based on Scoring guide, no evidence from the video given)

S2 : Level 2
(Explanation, no evidence given such as why N4 is an A- and not an A)

The structure in arguing against TF=9 is: at Level 2 – 2 occurrences.

Subsequently, B1 attempts to refute A1's argument that N4's performance merited only an A- for TF (Lines 81-89, Appendix F).

B1 argues that N4 deserves more than A- in TF (C1.6), seeking clarification and confirmation for the following:

- Referring to this one (S1)
 - o Participates and all that (S1.1)
 - o Keep discussion focus (S1.2)
 - o Lead and all that (S1.3)

The evaluation of the argument follows:

S1 and components: Level 4

(Linking TF criteria to skills both A1 and B1 had agreed, in a prior argument that N4 displayed. Secondhand evidence, hence not Level 5)

The structure of disputing N4's TF performance as A- is one occurrence of Level 4 argument.

A1 then follows the reference to TF criteria, citing the scoring guide and then offering that N4 belonged to the second band for TF in Lines 90-92 (refer to Appendix F for transcript). The analysis is below:

A1 reasons that N4 belongs to 2nd band for TF (C1.7) by stating:

- TF = 9: fulfills task competently, all tasks met with no difficulties (S1)
- I thought she was more of the second band (S2)
 - o (implicit: N4 had difficulties meeting some tasks)

The arguments are evaluated below:

S1 : Level 2
(Explanation based on scoring guide, no evidence from video)

S2 : Level 2
(Explanation or expression of personal opinion, no evidence from video to elaborate which tasks N4 had difficulties completing)

The structure of arguing that N4 belonged to the second band for TF is: at Level 2 – 2 occurrences.

The analysis provided above show that the argumentation for N4’s TF score comprises 4 claims and 6 arguments. The tabulation is presented in Table 4.29.

Table 4.29: Summary of the Level of Arguments by Pair 1 (TF)

Level of argument	A1	B1
Level 2	4	1
Level 3	-	-
Level 4	-	1
Level 5	-	-

4.4.1.3 Discussion on N4’s LC Score

Following the consideration of the justifications for and against their respective TF scores, the analysis will then proceed to the Language Control (LC) score. It is found that there is only one set of arguments where both A1 and B1 appear to collaborate their arguments in support of their respective standpoints as they appear to agree on the score banding that N4’s performance deserves.

As they move on to the discussion on N4’s LC score in Lines 115-123 (Appendix F), it appeared that both A1 and B1 were in agreement of N4’s performance from the perspective of language, both having assessed her in the middle score banding. The analysis can be found as follows:

A1 and B1 both agree that N4 deserves only a middle banding score for LC, (C1.8):

- She made some obvious language errors (S1 – B1)
- Yes [obvious language errors] (S1 – A1)
 - o Formulaic expressions (S1.1)
 - o Fossilized errors (S1.2)
 - o Typical Malaysian English’s missing ‘be’ (S1.3)
- A lot of fillers (S2 – B1)
- Agree [a lot of fillers] (S2 – A1)

The evaluation of B1’s arguments follow:

- S1 : Level 3
(Observation based on evidence from video, no examples or elaborations given)
- S2 : Level 3
(Evidence from video, no elaboration of quantity or how this affects the interaction)

The structure of B1’s arguments on N4 deserving the middle score banding for LC is: at Level 3 – 2 occurrences.

The evaluation of A1’s arguments follow:

- S1 : Level 4
(Expressed agreement with B1’s point on language errors, gave specific examples of the errors)
- S2 : Level 2
(Expressed agreement with B1’s point on fillers, no elaboration or examples given.)

The structure of A1’s arguments on N4 deserving the middle score banding for LC is: at Level 2 – 1 occurrence, at Level 4 – 1 occurrence.

The analysis of the argument involving N4’s LC score consists of 1 claim and 4 arguments from Pair 1 as they collaboratively justify N4’s performance at the middle score banding.

This is tabulated in Table 4.30.

Table 4.30: Summary of the Level of Arguments by Pair 1 (LC)

Level of argument	A1	B1
Level 2	1	-
Level 3	-	2
Level 4	1	-
Level 5	-	-

4.4.1.4 Discussion on N4's CA Score

The analysis will now move on to that of the score for N4's Communicative Ability (CA) where there are 8 sets of arguments:

- 1) A1 defending her standpoint,
- 2) B1's argument against A1's standpoint,
- 3) A1's argument for N4's fluency,
- 4) B1's argument against N4's fillers,
- 5) A1's arguments for N4's use of fillers,
- 6) B1's argument regarding N4's performance,
- 7) B1's argument against coming to a compromise on marks, and
- 8) A1's argument in favor of B1 raising her marks for N4.

A1 began the discussion on N4's CA score by highlighting how N4 had contributed to the group interaction, and presented these as communicative ability skills. A1 also addressed the two other criteria under CA as seen in Lines 163-175 (refer Appendix F for transcript). The analysis follows:

A1 justified that N4 deserves score of 7 for CA (C1.9), stating:

- She displayed CA skills/criteria (S1)
 - o She had reasons (S1.1)
 - o She contributed more to the discussion than the others, without dominating (S1.2)

- Knew what to say (S1.2.1)
- Knew how much to say (S1.2.2)
- Knew when to stop (S1.2.3)
- Knew when to bring others in (S1.2.4)
- Fluency – all of them had fillers (S2)
- Pronunciation errors present had not impeded comprehension (S3)
 - They all understood each other (S3.1)

The arguments are evaluated below:

S1 and components: Level 5

(Listed evidence from the video and linked these to the CA criteria, citing multiple examples)

S2 : Level 3

(Observation based on evidence from video, no elaboration on quantity or how it affects the interaction)

S3, S3.1 : Level 4

(Observation based on evidence from video, provided minimal support as elaboration)

The structure used in defending CA=7 is: at Level 3 – 1 occurrence, at Level 4 – 1 occurrence, and at Level 5 – 1 occurrence.

B1 then attempted to refute A1's standpoint (Lines 177-181, Appendix F), focusing on the hesitations in N4's performance, arguing that N4 does not deserve score of 7 for CA (C1.10), observing:

- Score category 7 for fluency: very few hesitations (S1)
 - (implicit: she had a lot of hesitations)
 - How was her performance [in terms of fluency] (S1.1)
 - (implicit: B1 is seeking A1's agreement on N4's numerous fillers)

The evaluation of the argument follows:

S1, S1.1 : Level 2

(Cited scoring guide, no evidence from video given, only implied. Made attempt to question A1's standpoint)

The structure used in refuting A1's standpoint is: at Level 2 – 1 occurrence.

Upon B1's questioning, A1 first establishes that they are discussing N4's CA and extended her impression of this score section before justifying her score by stating her observation of N4's performance in regards to hesitations. A1 then provided a scenario involving excessive hesitation before providing her (A1) impression of N4's hesitation. The analysis of Lines 183-198 (refer Appendix F for transcript) can be found below.

A1 argued that N4's hesitations do not affect fluency (C1.11), citing:

- I believe I marked her as having few hesitations (S1)
 - o Hesitations did not affect flow of arguments (S1.1)
 - Some people hesitate a lot, affecting message comprehension (S1.1.1)
 - o N4 had hesitations and fillers stem from searching for words (S1.2)
 - N4's hesitations not because of lack of ideas (S1.2.1)

The argument is evaluated below:

S1 and components: Level 5

(Explanation with evidence from video, illustrates point with scenario. Assessor linked impression of N4's performance with the scenario offered)

The structure used in arguing N4's performance in terms of hesitation is: at Level 5 – 1 occurrence.

Following A1's explanation involving N4's performance and hesitations, B1 attempted to argue that N4's numerous fillers construed hesitations (Lines 201-203, Appendix F).

The analysis is below:

B1 attempted to argue that Lots of fillers means a lot of hesitation (C1.12), but provides no support.

The evaluation of the argument follows:

C1.12 : Level 1
(No supporting argument given)

B1's argument on fillers being synonymous with hesitations is: at Level 1 – 1 occurrence

In Lines 204-216 (refer to Appendix F for transcript), A1 disputed the claim on N4's numerous fillers, arguing that fillers and hesitations were bad only in excess, and proceeded to expound on the uses of hesitation devices, further stating that she (A1) found that this was what N4 had done. She finished, stating it looked appropriate. The analysis is below:

N4 used fillers and hesitations appropriately (C1.13)

- N4 made effective use of hesitation devices (S1)
 - o Hesitated to find right word to say (S1.1)
 - o Hesitated before cutting into somebody's turn (S1.2)
- Fillers and hesitations are bad when excessive (S2)
 - o Excessive fillers/hesitation affects delivery of message (S2.1)
- N4's filler/hesitation use did not appear to be a cover for lack of proficiency (S3)

The arguments are evaluated below:

S1, S1.1 : Level 2
(Explanation and expression of personal opinion, no elaboration with evidence from video)

S2, S2.1 : Level 2
(Explanation, is not illustrated with evidence from video)

S3 : Level 2
(Explanation and expression of opinion, no evidence or example from video is provided)

The structure used in expressing N4's use of fillers is: at Level 2 – 3 occurrences.

B1 maintained opposition to A1's standpoint, pointing to N4's numerous fillers and language errors, A1 acknowledged the errors, noting an instance for which she had penalized N4. This is based on Lines 219-230 (refer to Appendix F for transcript). The analysis follows:

B1 maintained that N4 should not get score of 7 for CA (C1.14 – B1), stating:

- N4 had a lot of fillers (S1 – B1)
- N4 had a lot of language errors (S2 – B1)
- Agree, N4 had a lot of language errors (S1 – A1)
 - o Marked N4 down for saying 'reasonable' instead of 'reason' (S1.1 – A1)

B1's arguments are evaluated below:

S1 : Level 3
(Observation of evidence from video, no elaboration given)

S2 : Level 3
(Observation of evidence from video, no elaboration given)

The structure B1 used in opposing N4's standpoint is: at Level 3 – 2 occurrences.

A1's argument is evaluated below:

S1, S1.1 : Level 4
(Precise evidence or example from video used to illustrate argument)

The structure A1 used in supporting that N4 had many language errors is: at Level 4 – 1 occurrence.

After an aborted attempt to establish results, where B1 had rejected the offer of 6.5 marks, A1 had questioned if B1 wanted to review the group oral interaction, prompting B1 to defend her standpoint by citing N4's fluency, as can be seen in Line 265 (refer to Appendix F). The analysis is below:

B1 argues that N4 does not deserve more than score of 5 for CA (C1.15), she mentions:
- N4 was not very fluent (S1)

The evaluation of the argument follows:

S1 : Level 2
(Expression of opinion, no evidence from video given)

The structure used in defending the standpoint of CA=5 is: at Level 2 – 1 occurrence.

In Lines 266-278 (refer to Appendix F), A1 acknowledges B1's stance regarding N4's fluency, then went on to point out that there were three criteria within CA and that N4 deserved a higher mark for the other two criteria, and that fluency or hesitation was only one criterion. The analysis is as given as follows:

A1 counters, saying N4 deserves more than score of 5 for CA (C1.16), because:

- Fluency and hesitation is only one criteria in CA (S1)
 - o You may want to take down the marks for this (S1.1)
- There are two other criteria in CA (S2)
 - o Contributing to discussion satisfactorily (S2.1)
 - o Errors do not impede comprehension (S2.2)
 - We agreed earlier that message was comprehensible (S2.2.1)
 - o N4 deserves 6 or 7 for these two criteria (S2.3)

The arguments are evaluated below:

S1, S1.1 : Level 2

(Explanation and mention of rubric, no evidence from video was incorporated into argument)

S2 and components: Level 2

(Explanation and mention of rubric, reminder of what had been agreed upon earlier, no evidence from video was incorporated into argument)

The structure used in countering B1's standpoint is: at Level 2 – 2 occurrences.

From the analysis above, it can be seen that there are a total of 8 claims and 15 arguments exchanged by Pair 1 as they try to come to an agreement on the CA score for N4. This has been tabulated in Table 4.31.

Table 4.31: Summary of the Level of Arguments by Pair 1 (CA)

Level of argument	A1	B1
Level 1	-	1
Level 2	5	2
Level 3	1	2
Level 4	2	-
Level 5	2	-

4.4.2 Pair 2 (Assessors A2 and B2)

As was seen in the findings of the first and second research questions, the argumentation stage is present only in two of Pair 2's discussions, these being the discussion on N2's

Language Control (LC) score, and the discussion on N2's Communicative Ability (CA) score. However, due to space constraints, the arguments will be summarized here while the analysis can be found in Appendix J.

Pair 2 offers only one set of arguments for the discussion on N2's Language Control (LC) score. In this argument, A2 defends her standpoint of LC=8, citing a lack of vocabulary range and accuracy despite being more vocal and relatively fluent. A2 did allow that N2 had played the part of the leader and tried to get the discussion going, but added that more language errors surfaced as N2 spoke. A2 then invoked the scoring guide descriptions to support her argument. This is based on Lines 52-71 (refer to Appendix G for transcript). The analysis and evaluation of A2's argument to defend her standpoint can be found in Appendix J. The structure of A2's defense of her standpoint is: at Level 4 – 1 occurrence, at Level 5 – 1 occurrence.

The analysis indicate that there is 1 claim and 2 arguments where A2 defends her standpoint for N2's LC score. This is tabulated in Table 4.32.

Table 4.32: Summary of the Level of Arguments by Pair 2 (LC)

Level of argument	A2	B2
Level 1	-	-
Level 2	-	-
Level 3	-	-
Level 4	1	-
Level 5	1	-

Having reached an agreement on N2's Language Control (LC) score, Pair 2 then moved on to the discussion on N2's Communicative Ability (CA) score, which comprises 6 sets of arguments:

- 1) A2's brief explanation for the range of difference in their scores,
- 2) B2's rejection of A2's standpoint,
- 3) A2's defense of her standpoint,
- 4) A2's argument against B2's standpoint,
- 5) A2's argument against accepting B2's standpoint, and
- 6) A2's argument on N2's over leading.

Pair 2 begin their discussion on N2's CA score with A2 giving a brief explanation of the large difference between the CA scores which she and B2 had given N2. In Line 101-103 (refer to Appendix G for transcript), A2 attributes this difference to her being stricter in her marking. The analysis and evaluation of A2's explanation for the difference in their CA scores can be found in Appendix J. The structure of A2's explanation for the score difference is: at Level 1 – 1 occurrence.

In Lines 104-116 (refer to Appendix G for transcript), B2 subsequently rejects A2's explanation and standpoint, at the same time defending her own standpoint. B2 cites the scoring guide, pointing out the descriptor for the 8.5 marks she (B2) had given, and further stated the criteria for communicative ability which she compared to the skills N2 had displayed, including leadership skills. The analysis and evaluation of B2's arguments to reject A2's standpoint can be found in Appendix J. The structure of B2's rejection of A2's standpoint is: at Level 3 – 1 occurrence, at Level 5 – 1 occurrence.

A2 acknowledges B2's arguments for N2's ability to lead the discussion, then defended her own standpoint by explaining that she had not referred to the appendix listing the various CA skills, only having referred to the descriptors in the scoring guide. A2 then points out the difference in detail between the descriptors in the scoring guide and the skills listed in the CA appendix. This can be found in Lines 117-125 (refer to Appendix G for transcript). The analysis and evaluation of A2's argument in defending her standpoint is in Appendix J. The structure of A2's defense of her standpoint is: at Level 2 – 1 occurrence.

Following this, A2 then offered arguments against B2's standpoint in Lines 128-147 (refer to Appendix G for transcript). A2 argued that N2 did not merit 8.5 marks for CA, citing fluency issues. She pointed out that a score of 8.5 suggests little or no hesitation, which she argued, was not the case for N2 who hesitated a lot, and did not reflect the level of fluency expected for a score of 8.5. The analysis and evaluation of A2's arguments against B2's standpoint can be found in Appendix J. The structure of A2's argument against B2's standpoint is: at Level 4 – 1 occurrence.

A2 had voiced consideration for withdrawing her standpoint, prompting B2 to consider withdrawing her own (B2) standpoint. This resulted in an attempt to resolve their difference, and A2 reiterates her explanation for not having referred to the CA appendix. Then A2 acknowledges that N2 had displayed relevant CA skills before arguing that N2 had over led, depriving others in the group of a chance to participate. A2 then observed that the rating scale makes no allowance for this, to which B2 agrees and points out that the group merely agreed with one another. A2 acknowledges this. This can be found in Lines 159-177 (refer Appendix G for transcript). The analysis and evaluation is in Appendix J. The structure of A2's argument against accepting B2's standpoint is: at Level

3 – 2 occurrences, at Level 4 – 1 occurrence, while the structure of B2’s consideration of CA criteria in regard to over leading is: at Level 3 – 1 occurrence.

Following this, in Lines 178-200 (refer to Appendix G for transcript), upon B2’s prompting, A2 extended her arguments on N2’s over leading. A2 argued that N2 took on the leadership role to talk and had not invited participation, at all, from the others in the group, especially the weaker students. B2 attempted to counter argue, then agreed that N2 had not invited participation from the weaker students. A2 then added that N2 had held a dialogue with N4, garnering protest from B2 that N4 had invited participation, prompting A2 to redirect them back to N2 before acknowledging that N2 had kept the discussion going. The analysis and evaluation for this argument can be found in Appendix J. The structure of A2’s argument regarding N2’s over leading is: at Level 3 – 2 occurrences, at Level 4 – 1 occurrence while the structure of B2’s counter arguments on the topic is: at Level 1 – 1 occurrence, at Level 2 – 1 occurrence, at Level 3 – 1 occurrence.

Pair 2 was found to have offered a total of 6 claims and 15 arguments as they tried to reach a consensus on the scores for N2’s Communicative Ability (CA) score. This is tabulated in Table 4.33.

Table 4.33: Summary of the Level of Arguments by Pair 2 (CA)

Level of argument	A2	B2
Level 1	1	1
Level 2	1	1
Level 3	4	3
Level 4	3	-
Level 5	-	1

4.4.3 Comparison between Pair 1 and Pair 2

Table 4.34: Comparison of arguments and Levels of Arguments of Pairs 1 and 2

Discussions	PAIR 1		PAIR 2	
		4		2
Claims	16		7	
Arguments	35		17	
Level of argument	A1	B1	A2	B2
Level 1	-	1	1	1
Level 2	11	4	1	1
Level 3	1	8	4	3
Level 4	4	3	4	-
Level 5	3	-	1	1
TOTAL	19	16	11	6

As can be seen in Table 4.34, argumentation stages were present in all four of Pair 1's discussions under analysis while only two of Pair 2's discussions under analysis have argumentation stages. As a result, Pair 1 appear to have more than twice the number of claims that Pair 2 do, and following that, a little over twice the number of arguments or supports. It is apparent that both Pairs 1 and 2 have arguments ranging from Level 1 to Level 5 as they present justifications for the judgement they had rendered. While A1 and B1 appear to put forth nearly the same number of arguments in total, A2 appears to have made nearly twice the number of arguments B2 has. The bulk of the arguments put forth by both Pairs 1 and 2 appear to be Level 3 and above, suggesting that most of the arguments are based on evidence from the assessed group oral interaction. However, the high number of Level 2 arguments by Pair 1, 15 to Pair 2's 2, appear to show a tendency on the part of Pair 1, A1 especially, to use explanations in justifying her scores, as was seen in the analysis. This high frequency of Level 2 arguments coupled with the large number of evidence based arguments suggests, perhaps, a depth to Pair 1's argumentation compared to that of Pair 2.

4.5 Conclusion

This chapter covered the findings of the analysis done on the data obtained from the study. It has presented the findings from the following: 1) the analysis involving a comparison of the data and the Ideal System for Resolution of Disputes, 2) the analysis of the moves and speech acts made by the assessors in the argumentation stages, and 3) the analysis and evaluation of the arguments put forth by the assessors in justifying their scores. The next chapter will briefly summarize and discuss the findings here before concluding the study.

University of Malaya

CHAPTER 5

DISCUSSION OF FINDINGS AND CONCLUSION

5.1 Introduction

This chapter will discuss the findings presented in Chapter 4. Sections 5.2 to 5.4 will summarize the findings of each of the three research questions respectively and relate the current findings to other research done in a similar context. Following that, Section 5.5 will share and discuss some general findings deemed relevant to the study but was not part of the scope analyzed in the three research questions guiding this study. The chapter then touches on the limitations and possible avenues of future research and finally concludes with the implications that may be drawn from the findings and the possible impact of these implications on the role of moderation.

5.2 Ideal System for Resolution of Disputes

Both similarities and differences were found when the data was compared against the Ideal System for Resolution of Disputes. In terms of similarities with the ideal system, it was found that all of the confrontation stages and argumentation stages found in the analysis conform fully to what Van Eemeren et al. (1993) had suggested in the Ideal System for Resolution of Disputes.

While the data shared similarities with the ideal system, the differences found between the two were much more notable. The most notable of these differences were that the concluding stages found in both Pair 1 and Pair 2's moderation which did not follow the model found in the ideal system. It was found that in all the discussions analyzed for Pair

1 and Pair 2, the moderating assessors did not achieve resolution but settled their differences. It is noted, however, that despite the tendency of all four moderating pairs resorting to settlement to reach a consensus on scores, one instance of resolution was found, this was in Pair 3's moderation, specifically in the discussion on Language Control (LC) scores where A3 accepted B3's standpoint (Refer Appendix K).

According to Van Eemeren, Grootendorst, and Snoeck Henkemans (2002), a resolution is achieved when the standpoint which did not stand up to criticism is retracted and the participant whose standpoint is retracted accepts the standpoint of their opponent. With the exception of Pair 3's discussion on LC scores, it was found instead that in all of the concluding stages involving a numerical standpoint, both participants in each pair revised or retracted their standpoints, resulting in a "settlement" as termed by Van Eemeren, Grootendorst, and Snoeck Henkemans (2002). They define "settlement" as an occurrence where a result has been reached and the discussion ended but no resolution and no agreement was achieved, and therefore the difference was not resolved only put aside and may recur.

Another notable difference is that the discussions did not necessarily follow the order proposed in the ideal system where it was found that some stages of discussion were skipped but occurred later or out of order, or that certain stages of discussion were revisited one or more times. This holds true not just for the discussions found in Pairs 1 and 2 but in the discussions of Pairs 3 and 4 as well. While there does not appear to be literature addressing this, this did not seem to have a particular impact on the success of a critical discussion provided the discussion concludes with the concluding stage as proposed in the ideal system where a resolution is achieved.

As a whole, it becomes apparent that the ideal system is a guideline for the analysis and evaluation of a discussion, any discussion, critical or otherwise. Finding similarities and differences between actual spontaneous discussion and the discussion model suggested in the Ideal System for Resolution of Disputes apparently does not aid in the determination if the discussion is effective. However, by using the ideal system to identify the different parts or discussion stages in a particular discussion, it may allow the analyst to focus on various elements within those discussion stages, and from there stand a chance to examine what the strength and weaknesses of the discussion may be as is found and demonstrated by Van Rees (2003).

5.3 Moves and Speech Acts Used to Justify Scores Given

For the second research question, it was found that the 'A' assessors use more speech acts than the 'B' assessors, assertives in particular while the number of commissives used by the paired moderators appear to be similar in number. This finding from the analysis of Pair 1 and Pair 2's moderation was also present in the moderations of Pair 3 and Pair 4. The notable finding was that B1 uses twice the number of directives that A1 does while Pair 2 does not use any directives. An almost similar trend of directive use was found in Pairs 3 and 4, where A3 and B3 both used one directive each, and B4 used two directives while A4 none. It appears that the greater number of assertives used by the 'A' assessors is due to the greater occurrence of the following moves: putting forth arguments, making explanations, and making statements. It could perhaps be attributed to the experience variable, since the 'A' assessors possess greater experience in both teaching and assessing group oral interaction than the 'B' assessors and may therefore be more authoritative than the 'B' assessors when it comes to assessment and particularly in justifying or defending the scores they had given. This appears in line with the findings of some studies cited by

Orr (2007) where she indicates that seniority is or could be an influencing factor in the outcome of a moderation discussion. It was also one of Orr's finding in her study to illustrate the differing natures between the positivist approach to moderation in theory and the poststructuralist approach found in actual moderation discourse.

It may also be noted that although Van Eemeren and Grootendorst (2004) have stated that expressives do not have a proper role in the pragma-dialectical study of argumentation, it was found that B1 had made use of one instance of an expressive in one of her arguments. It did not appear to play a part other than to express why she did not wish to accept a point A1 had made and perhaps had not influenced A1, contrary to Van Eemeren and Grootendorst's (2004) suggestion that an expressive may influence the outcome of a critical discussion.

The findings for this question suggest that experience may have fostered greater confidence in an assessor's perception of their ability when it comes to test assessment, where it appears that they are more willing to put forth arguments and make explanations and statements of what they may have observed of a test. It also seems that the number of explanations made by 'A' and 'B' assessors respectively suggest that the 'A' assessors treat the moderation session as a chance to share knowledge and experience. This concurs with Wyatt-Smith, Klenowski, and Gunn's (2010) assertion that moderation is a chance for assessors to share and build interpretations and knowledge. This may explain B1 and B4's use of directives to ask questions and seek clarification, thus suggesting that B1 and perhaps by extension other less experienced assessors, treat moderation as a chance to learn from their more experienced counterparts.

5.4 Structure and Presentation of Justifications

It was found that the 'A' assessors put forth more arguments than their 'B' counterparts, and in addition, appear to have a tendency to use more higher level arguments (Level 4 and Level 5 arguments). Again, this could be attributed to the experience variable as the 'A' assessors hold seniority over the 'B' assessors. It might be concluded that the greater experience the 'A' assessors possess, the higher their ability of noticing more detailed observations in an assessment, and this may have allowed them be better able to explain what they may have observed and perhaps even connect criteria and qualities to a candidate's performance. This may tie back to the seniority factor Orr (2007) found and cited from other studies. Further, Wyatt-Smith, Klenowski, and Gunn's (2010) findings revealed that moderation may be an avenue in which teachers come together to "share interpretations of criteria and standards" (p. 61), and this appears to be in line with the findings of the current study. Adie, Klenowski, and Wyatt-Smith (2012) who cites shared understanding of terminology as a notable factor in moderation also leads to a similar understanding that having greater experience and seniority, the 'A' assessors apparently have a better grasp of the interpretations of criteria and terminology, resulting in them having more confidence in their judgement. Additionally, Wyatt-Smith, Kenowski, and Gunn (2010) found that teachers value fairness and may assess on a 'deserve' or 'merit' basis in favor of criteria. Following this, the 'A' assessors may perhaps have drawn on prior performances they have assessed and put forth arguments from that front, as may be suggested by A1's arguments regarding hesitations and fillers in the discussion on N4's CA score.

Not only do the more experienced assessors appear to perhaps have an ability of noticing more detailed evidence in an group oral interaction test where there is usually no chance

of reviewing the assessment piece, the experience factor may also have taught or trained the assessors what they have to focus on in such test situations. And having more experience in assessment and post-assessment moderation may have taught them the importance of being able to identify the reasons underlying the scores they had awarded. Additionally, this experience may have imparted on them the importance of being able to explain, defend and justify the rationale behind the marks they had given.

5.5 Overview of the Findings

This section will discuss some general findings not addressed in the findings and analysis of the three research questions. The study has provided some valuable insights to how assessors put forward their arguments when they score students in a group oral interaction test situation. It appears that the overall outcome favors the 'A' assessors as evident from the outcomes of Pair 1's discussions of N4's TF and CA scores as well as Pair 2's discussion on N2's CA score. This trend is also present in Pair 3's discussion on N1's CA score and in Pair 4's discussion on N1's LC score (refer Appendix K). While the 8 marks Pair 1 agreed on for N4's TF mark appears to be midway between A1's 7 and B1's 9 marks, it is noted that that in the rating scale, for the category of TF, 7-8 is one score band while 9-10 is another score band. N4's CA score outcome is more obvious as the pair agreed on a final score of 6.5 against A1's 7 and B1's 5 marks. Again with Pair 2, for N2's CA score, the pair agreed on a score of 6.5 against A2's 5 and B2's 8.5 marks. However, for Pair 2's discussion on N2's LC score, the balance swung in the opposite direction as seen by the agreed final score of 9.5 against A2's 8 against B2's 10.5 marks.

Another instance where the outcome favored the 'B' assessor was in Pair 3's discussion on LC score where the difference was resolved when A3 accepted B3's standpoint (Refer

Appendix K). This tendency to adjust the mark in favor of one assessor appears to be one of the ways in which moderating assessors deal with a disagreement which Orr (2007) attributed to Dracup's (1997) "the agreement process". This tendency for the outcomes to favor the more experienced assessor concurs with Orr's observation of the role of power relations in moderation where the more experienced assessors were less likely to give in in the case of a difference in scores. According to Orr, yet another agreement tactic when there is a disagreement in scores is the by splitting the difference. This was also found in the current study, evident in Pair 1's discussion on N4's TF and LC scores, and suggested in Pair 2's discussion on N2's TF score. This leads to a tentative conclusion that these are common and accepted moderation practices.

In addition to some specific observations of the findings as given above, it also appears that despite the sometimes large difference in scores between the assessors, in their attempts to reach a consensus through discussion, often times, they mention details from the group oral interaction they had assessed. And more importantly, it is observed that the moderating assessors refer to the scoring guide as they attempt to reach said consensus. Both these points can be found in the findings for the third research question. This shows that that all the participating assessors have high regard for the rating criteria, as they could have consciously or unconsciously recognized its importance in assessment and moderation. This appears to support Van der Schaaf, Baartman, and Prins's (2012) findings on the role of criteria in assessment moderation. It suggests that the participants of this study, while perhaps not the most consistent of assessors, do accept the significant role of moderation, which is to ensure fairness and reliability, hence turning to the rating criteria as a means of achieving this goal of moderation.

5.6 Limitations and Avenues for Future Study

It is to be noted that this study is rather limited in scope as it only actively analyzes various short segments each of two paired moderation sessions with another two pairs used to corroborate what was found in the first two. All the participating assessors were drawn from the academic teaching staff of one faculty where the tests are all conducted within the department, hence these findings cannot be generalized to a larger context. In addition, most of the assessors have received little or no training in the assessment of group oral interaction despite being assessors of such test formats. Besides, most of the assessors were not familiar with the course and rating scale used in the study, and it is quite likely they were not familiar with the standards and interpretations of this particular rating scale.

Having said that, the discussions found in the moderation discourse have provided some enlightening insights to the whole moderation practice and one possible avenue for future study could look into the power relations between the moderating pairs.

5.7 Conclusion

This section will discuss what implications the findings may suggest for moderation and assessment. This study has shown the role moderation plays in fostering confidence in the assessment process. This is seen in how each pair of assessors make the effort to discuss the marks they had given by either defending their score, or arguing against the score their opponent had given to the students in the video-taped group oral interaction that they had each assessed individually prior to the moderation. Although the Van Eemeren and Grootendorst's (2004) pragma-dialectical framework and the Ideal System for Resolution of Disputes appears to not be entirely suitable for the analysis of moderation discourse, it is observed that the framework and the ideal system has helped to narrow the focus, thus allowing for more detailed analysis, and through that leading to an understanding of the discussion process which occurs in moderation. For it is through the discussion stages of the ideal system that allows the analyst to concentrate on crucial segments of the discussion, aiding in the analysis process. This then allows for the identification of the moves the moderating assessors engage in during the argumentation stage, and following that the evaluation of the relevance, structure and depth of the arguments put forth. Thus allowing for the identification and determination of what happens in the process of reaching a consensus in moderation.

The findings of the first research question show that for each of the moderating pairs, oftentimes, consensus is achieved not through the pragma-dialectical resolution, but through settlement where both participants of each moderating pair compromise their differences. These compromises were reached through splitting the difference, whether fairly or in favor of one party, thus supporting Orr's (2007) findings. The findings of the second and third research questions also parallel the findings of Orr's (2007) study, in

which she notes that moderation when practiced is poststructuralist in nature where consensus is achieved through an exchange of rhetorical moves. And is, perhaps, the lead up to the decision to compromise on the scores.

The discussions found in the moderations analyzed in this study show that the assessors appear to take the moderation sessions seriously in that they discuss details critically. Observations and evidence from the video-taped group oral interaction were taken into consideration as they acknowledged the students' merits, achievements, performance and shortcomings, and compared these against their interpretations of the scoring guide. In the process, they offered arguments, defenses and justifications regarding the score they each believe the student deserves or does not deserve. Although it appears reassuring that the moderating assessors discuss, defend, justify, and argue the scores they had given to students in the group oral interaction to reach a compromise both parties appear to agree on. It also raises questions and implications for the role of moderation as a means of ensuring reliability. For the terms settlement and compromise suggest a less desirable connotation that a more accurate, deserving or fair result was not reached, suggesting that moderation is perhaps then not fair or reliable nor accountable.

This implication on the possibility of moderation not being as reliable as intended is compounded by the various factors that appear to influence the outcome of a moderation discussion. The findings of this study appear to indicate experience as the influencing factor. This is seen in how the more experienced assessor in each pair use more assertives during the discussion and make a greater number of higher level arguments. This difference in experience coupled with more persuasive rhetoric does perhaps already influence the score outcome. Additionally, Orr (2007) has cited studies stating that power relations also play a role in moderation. Also, Van der Schaaf, Baartman, and Prins (2012)

found that there appears to be a camaraderie in moderation where they complement what each other say, extend agreement easily, and have a tendency to influence one other. All this paints what may perhaps be a negative picture of the nature of moderation in contrast to its actual intended purpose of ensuring reliability and enforcing accountability. However, despite all the apparent shortcomings which may have surfaced regarding moderation practice, the fact that it brings together two or more assessors to discuss the scores for an assessment is a positive point as more than one opinion on a student's performance is discussed. This is especially crucial in performative assessments such as oral tests and particularly tests of group oral interaction given the fleeting nature of such tests and the myriad of factors and details to keep track of. Additionally, that all the moderating assessors do refer to the rating criteria as they discuss their difference in scores appear to suggest a recognizance of the role of moderation as a means of ensuring reliability and fairness in assessment, and it appears that they have made efforts to achieve this lofty goal. Hence, while not as reliable as theoretically implied, moderation still has a role to play in terms of reliability in assessment situations and will possibly only grow in effectiveness as it gains in popularity as a part of the assessment process.

REFERENCES

- Aakus, M. (2003). Neither Naïve nor Critical Reconstruction: Dispute Mediators, Impasse, and the Design of Argumentation. *Argumentation* 17: 265-290.
- Adie, L., Klenowski, V., and Wyatt-Smith, C. (2012). Towards an understanding of teacher judgement in the context of social moderation. *Educational Review*, 64: 223-240.
- Adie, L., Lloyd, M., and Beutel, D. (2013). Identifying Discourses of Moderation in Higher Education. *Assessment & Evaluation in Higher Education*, 38: 968–977.
- Baartman, L., Prins, F., Kirschner, P, and Van der Vleuten, C. (2007). Determining the quality of Competence Assessment Programs: A self-evaluation procedure. *Studies in Educational Evaluation*, 33: 258-281.
- Barrie, S., Hughes, C., and Crisp, G. (2014). *Assessing and Assuring Graduate Learning Outcomes (AAGLO) Final Report*. Office for Learning and Teaching: Sydney. Retrieved from <http://www.itl.usyd.edu.au/projects/aaglo/overview.htm>
- Bizup, J. (2009). The uses of Toulmin in Composition Studies. *Conference on College Composition and Communication*, 61: W1-W23.
- Bloxham, S. (2009). Marking and moderation in the UK: false assumptions and wasted resources. *Assessment and Evaluation in Higher Education*, 34: 209-220.
- Bloxham, S., Hughes, C., & Adie, L. (2015). What's the point of moderation: A discussion of the purposes achieved through contemporary moderation practices. *Assessment and Evaluation in Higher Education*. (In Press).
- Boud, D. (1995). Assessment and learning: contradictory or complementary? In ed. P. Knight. *Assessment for Learning in Higher Education*. London: Kogan Page.
- Clark, D.B., and Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education*, 29: 253–277.
- Cutting, J. (2008). *Pragmatics and Discourse: A resource book for students*. 2nd Ed. London: Routledge.
- Driver, R., Newton, P., and Osbourne, J. (2000). Establishing the Norms of Scientific Argumentation in Classrooms. *Science Education*, 84: 287-312.
- Ducasse, A. M. and Brown, A. (2009). Assessing paired orals: raters' orientations to interaction. *Language Testing*, 26: 423-443.

- Erduran, S., Simon, S., and Osborne, J. (2004). Tapping into argumentation: Developments in application of Toulmin's argument pattern for studying science discourse. *Science Education*, 88: 915–933.
- Finegan, E. (2008). *Language: Its Structure and Use (5th Ed.)*. Thomson Wadsworth.
- Fromkin, V., Rodman, R., & Hyams, N. (2011). *An introduction to language (9th ed.)*. USA: Wadsworth Cengage Learning.
- Huang, Y. (2007). *Pragmatics*. Oxford University Press.
- Johnson, G. M. (2015). Record of Assessment Moderation Practice (RAMP): Survey Software as a Mechanism of Continuous Quality Improvement. *Assessment & Evaluation in Higher Education*, 40 (2): 265-278.
- Jorgensen, M. and Philips, L. J. (2002). *Discourse Analysis as Theory and Method*. London: Sage Publications.
- Kaptein, H. (2005). Legal Progress Through Pragma-dialectics? Prospects Beyond Analogy and E Contario. *Argumentation*, 19: 497-507.
- Kohoutek, J. (2014). European Standards for Quality Assurance and Institutional Practices of Student Assessment in the UK, the Netherlands, and the Czech Republic. *Assessment & Evaluation in Higher Education*, 39 (3): 310-325.
- Lauerbach, G. (2007). Argumentation in political talk show interviews. *Journal of Pragmatics*, 39: 1388-1419.
- Lewnski, M. (2010). *Internet political discussion forums as an argumentative activity type: A pragma-dialectical analysis of online forms of strategic maneuvering in reacting critically*. Doctoral Dissertation: University of Amsterdam. Rozenberg Publications.
- McNamara, T. (1996). *Measuring Second Language Performance*. Longman.
- McNamara, T. (2000). *Language Testing*. Oxford University Press.
- Mohammed, D. (2006). *Towards a Pragma-Dialectical Approach to Negotiation*. Paper presented at the Sixth Conference of the International Society for the Study of Argumentation.
- Orr, S. (2007). Assessment Moderation: Constructing the Marks and Constructing the Students. *Assessment & Evaluation in Higher Education*, 32: 645–656.

- Orr, S. and Bloxham, S. (2013). Making Judgements about Students Making Work: Lecturers' Assessment in Art and Design. *Arts and Humanities in Higher Education*, 12: 234-253.
- Pedersen, S. (2010). *Reasonable Non-Agreement in Critical Discussions*. Paper presented at the Seventh Conference of the International Society for the Study of Argumentation in Amsterdam.
- Reiman, N., Harman, K., Wilson, A., & McDowell, L. (2010). *Learning to Assess in Higher Education*. Paper presented at the Higher Education Close Up Conference in Lancaster.
- Robinson, D. (2006). *Introducing Performative Pragmatics*. Routledge.
- Sadler, D. R. (2013). Assuring Academic Achievement Standards: From Moderation to Calibration. *Assessment in Education: Principles, Policy and Practice*, 20 (1): 5-19.
- Thirlwall, A. (2008). *Analyzing problem-solving in negotiations: A pragmatic approach*. Masters research thesis: University of Amsterdam.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Tracy, K. & Mirivel, J. (2009). Discourse Analysis: The practice and practical value of taping, transcribing, and analyzing talk. In L. R. Frey & K. N. Cissna (Eds.), *Handbook of Applied Communication Research*, 153-179. New York: Routledge.
- Van der Schaaf, M., Baartman, L., Prins, F. (2012). Exploring the role of assessment criteria during teachers' collaborative judgment process of students' portfolios. *Assessment and Evaluation in Higher Education*, 37: 847-860.
- Van Eemeren, F. H. and Grootendorst, R. (1984). *Speech acts in argumentative discussions: A theoretical model for the analysis of discussions directed towards solving conflicts of opinions*. Dordrecht: Floris Publications.
- Van Eemeren, F. H., Grootendorst, R., Jackson, S., and Jacobs, S. (1993). *Reconstructing Argumentative Discourse*. The University of Alabama Press.
- Van Eemeren, F., Grootendorst, R. and Snoeck Henkemans, A. F. (2002). *Argumentation: Analysis, Evaluation, Presentation*. L. Erlbaum Associates.
- Van Eemeren, F. and Houtlosser, P. (2002). Strategic maneuvering with the burden of proof. In Van Eemeren, F. (Ed.). *Advances in pragma-dialectics*. Amsterdam: SicSat.

- Van Eemeren, F. and Houtlosser, P. (2006). Strategic Maneuvering: A synthetic recapitulation. *Argumentation*, 20: 381-392.
- Van Eemeren, F. and Grootendorst, R. (2004). *A Systematic Theory of Argumentation: The pragma-dialectical approach*. Cambridge University Press.
- Van Rees, M. A. (2003). Pragma-Dialectical Analysis and Evaluation of Problem-Solving Discussion. *Argumentation*, 17: 465-479.
- Wyatt-Smith, C., Klenowski, V., and Gunn, S. (2010). The centrality of teachers' judgment practice in assessment: a study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17: 59-75.
- Yorke, M. (2008). *Grading Student Achievement in Higher Education: Signals and Shortcomings*. Abingdon: Routledge.
- Yule, G. (1996). *Pragmatics*. Oxford University Press.
- Zarefsky, D. (2006). Strategic Maneuvering through Persuasive Definitions: Implications for Dialectic and Rhetoric. *Argumentation*, 20: 399-416.