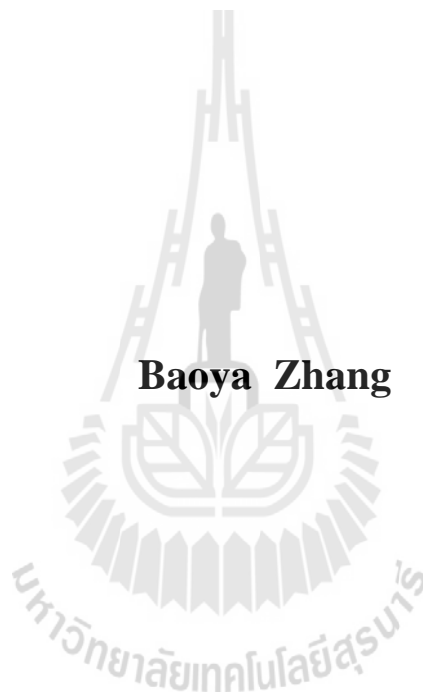# MOVES AND INTER-MOVE LINGUISTIC VARIATION

IN EDUCATION RESEARCH ARTICLES

Baoya Zhang

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy in English Language Studies

Suranaree University of Technology

Academic Year 2015

อัธภาคและความแตกต่างทางภาษาศาสตร์ระหว่างอัธภาค
ของบทความวิจัยทางการศึกษา

นายเป่าหย้า จาง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต
สาขาวิชาภาษาอังกฤษศึกษา
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2558

MOVES AND INTER-MOVE LINGUISTIC VARIATION IN EDUCATION RESEARCH ARTICLES

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

_____

(Asst. Prof. Dr. Issra Pramoolsook)

Chairperson

_____

(Assoc. Prof. Dr. Anchalee Wannaruk)

Member (Thesis Advisor)

_____

(Assoc. Prof. Dr. Kanyarat Getkham)

Member

_____

(Asst. Prof. Dr. Apisak Pupipat)

Member

_____

(Asst. Prof. Dr. Darunee Dujsik)

Member

_____   _____

(Prof. Dr. Sukit Limpijumnong)   (Dr. Peerasak Siriyothin)

Vice Rector for Academic Affairs   Dean of Institute of Social Technology

and Innovation

เป่าหย้า จาง : อัฒภาคและความแตกต่างทางภาษาศาสตร์ระหว่างอัฒภาคของบทความวิจัย
ทางการศึกษา (MOVES AND INTER-MOVE LINGUISTIC VARIATION IN
EDUCATION RESEARCH ARTICLES) อาจารย์ที่ปรึกษา :
รองศาสตราจารย์ ดร.อัญชลี วรรณรักษ์, 304 หน้า

งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อศึกษาประเภทการสื่อสาร (genre) ของบทความวิจัย
ทางด้านศึกษาศาสตร์โดยใช้รูปแบบผสมสองขั้นตอน เพื่อศึกษารายละเอียดของประเภทการสื่อสาร
ทางด้านโครงสร้างอัตถภาคเชิงวาทศิลป์ (rhetorical move) และความหลากหลายทางภาษาศาสตร์
ระหว่างอัตถภาคโดยมีวัตถุประสงค์หลักสองประการคือ 1) เพื่อวิเคราะห์อัตถภาคเชิงวาทศิลป์และ
ลำดับของอัตถภาคของบทความวิจัยด้านศึกษาศาสตร์และ 2) เพื่อวิเคราะห์รูปแบบการปรากฏ
ร่วมกันของลักษณะทางภาษา รวมทั้งความเหมือนและความต่างของรูปแบบของอัตถภาคที่ปรากฏ
ร่วมกัน ผู้วิจัยสร้างคลังข้อมูลอีอาร์ซี (ERC) ซึ่งประกอบด้วยบทความวิจัยทางด้านการศึกษา 120
บทความ ในช่วงแรกของงานวิจัยเป็นการวิเคราะห์ประเภทการสื่อสารตามแนวอัตถภาควิเคราะห์
ของ Swales' (1990, 2004) เพื่อวิเคราะห์โครงสร้างของอัตถภาคเชิงวาทศิลป์ หลังจากนั้นผู้วิจัยทำ
การวิเคราะห์หลายมิติตามแนวทางของ Biber's (1986, 1988) เพื่อวิเคราะห์ลักษณะทางภาษาที่มัก
ปรากฏร่วมกันบ่อยครั้งในแต่ละอัตถภาค ผลการวิจัยพบอัตถภาค 16 ประเภทพร้อมลักษณะการ
เรียงลำดับของอัตถภาค จำแนกตามส่วนต่างๆของบทความวิจัย (บทนำ วิธีวิจัย ผลการวิจัยและ
อภิปรายผลการวิจัย) สำหรับช่วงที่สองของงานวิจัยเป็นการ วิเคราะห์หลายมิติโดยพบลักษณะของ
การสื่อสารพื้นฐานเจ็ดประเภทกล่าวคือ 1) *Current information vs. procedural concerns* 2)
Evaluative stance vs. past actions or states 3) *Logical probability vs. integrated information* 4)
Commentary 5) *Personal engagement vs. modified information* 6) *Unsatisfactory status quo vs.
research conduct* และ 7) *References to present research vs. Information relevant to the past* จาก
การวิเคราะห์หลายมิติพบว่าอัตถภาคมีทั้งส่วนที่เหมือนและต่างกัน ผลการวิจัยทำให้เข้าใจประเภท
การสื่อสารของบทความวิจัยทางด้านการศึกษาศาสตร์มากขึ้น การประยุกต์ใช้ผลการวิจัยทำได้โดย
การศึกษาและยึดปฏิบัติตามรูปแบบโครงสร้างและรายละเอียดทางภาษา นอกจากนี้นักวิจัยมือใหม่
ทางด้านการศึกษาทั้งที่เป็นเจ้าของภาษาและ ไม่ใช่เจ้าของภาษาสามารถใช้ผลการวิจัยนี้เป็นแนวทาง
ในการเขียนบทความวิจัยเพื่อตีพิมพ์ในระดับนานาชาติได้ ผลการวิจัยสามารถใช้เป็นแนวทางใน
การออกแบบหลักสูตรที่เกี่ยวข้องกับการ ใช้ภาษาอังกฤษเฉพาะกิจ (ESP) หรือภาษาอังกฤษเพื่อ
วิชาการ (EAP) การพัฒนาสื่อการสอนเพื่อช่วยปรับปรุงการเรียนการสอนให้ดีขึ้น

สาขาวิชาภาษาต่างประเทศ   ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2558   ลายมือชื่ออาจารย์ที่ปรึกษา_____

BAOYA ZHANG : MOVES AND INTER-MOVE LINGUISTIC

VARIATION IN EDUCATION RESEARCH ARTICLES.   THESIS

ADVISOR : ASSOC. PROF. ANCHALEE WANNARUK, Ph.D.,

304 PP.

EDUCATION RESEARCH ARTICLES/ CORPUS/ RHETORICAL MOVE/ MOVE

ANALYSIS/ DIMENSIONS/ MULTIDIMENSIONAL ANALYSIS

The present study investigated the education research article genre using a two-phase sequential mixed methods design. The purpose was to achieve a comprehensive description of the genre with regards to the rhetorical move structure and linguistic variation among the moves. Specifically, two objectives were attempted: 1) to identify the rhetorical moves and their sequencing in education research articles, and 2) to identify the co-occurrence patterns of linguistic features in the articles and the similarities and differences among the moves with respect to the use of the co-occurrence patterns. A corpus called the ERC was constructed of 120 systematically sampled empirical education research articles. In the first phase, genre analysis within the framework of Swales' (1990, 2004) move analysis was conducted on the corpus to identify the rhetorical move structure. In the second, the corpus was subjected to Biber's (1986, 1988) multidimensional (MD) analysis to identify the linguistic features co-occurring with high frequencies to realize those rhetorical moves. The move analysis

identified 16 rhetorical moves in the corpus, and with their preferred order determined, a generic move structure was proposed for each IMRD section (Introduction, Methods, Results, and Discussion). In the second phase, the MD analysis uncovered seven basic functional dimensions of the education research article genre: *Current information vs. procedural concerns*, *Evaluative stance vs. past actions or states*, *Logical probability vs. integrated information*, *Commentary*, *Personal engagement vs. modified information*, *Unsatisfactory status quo vs. research conduct*, and *References to present research vs. Information relevant to the past.* The MD analysis revealed that moves of the same section may be quite similar or significantly different in the extent to which they are marked by the dimensions, and that moves of different sections may be similar on certain dimensions but significantly different on others. Overall, the results provide a comprehensive insight into the education research article genre. The practical relevance of the study is clear: by studying and following the generic structural models and linguistic descriptions, both novice NS and NNS education researchers can write better research articles for international publication. Pedagogically, these results can inform ESP or EAP syllabus design, materials development and improve classroom instruction.

School of Foreign Languages   Student's Signature _____

Academic Year 2015   Advisor's Signature _____

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

CARS   Create a Research Space

EAP   English for Academic Purposes

ERC   Education Research Corpus

ERC-M   Education Research Corpus-Moves

ESP   English for Specific Purposes

IMRD   Introduction, Methods, Results, Discussion

MD   Multidimensional

NNS   Non-native speakers

NS   Native speakers

POS   Part of Speech Tagging

RA   Research article

CHAPTER 1

INTRODUCTION

This introductory chapter presents an overview of a corpus-based study of the rhetorical move structure and inter-move linguistic variation of education research articles (RAs). While Section 1.1 advances the research problem to be addressed, Section 1.2 outlines the theoretical and empirical background of the investigation. Section 1.3 previews the study by outlining its objectives and research questions, and Section 1.4 outlines the significance of the research. The chapter concludes with a list of defined terms that are either important or specific to the study.

1.1 Research Problem

Education research is a vital resource for improving educational policies, programs and classroom instruction (Bassey, 1999; Mosteller, Nave, & Miech, 2004). Its value consists in its immediate relevance to its consumers and is attested by the sheer bulk of research articles (RAs) published in the field every year. For example, the number of journals of education research currently indexed by ERIC (Education Resources Information Center) totals 967 (United States Department of Education, 2015). Nevertheless, despite the potential value of education research, "how to move research findings into the field" effectively to benefit further education research and

practice is a complex and challenging problem. This problem is related to the diversity of education research, researchers and research consumers, but in a technical sense, it also concerns how quality research can be written up and accepted by the educational community.

Research articles (RAs) are written by researchers who wish to communicate new knowledge to the academic community and convince them of their research findings (Hyland, 2000). As an important academic genre, the RA has received much attention from genre analysts over the past decades, who have approached it from either a macro or micro perspective. Findings from a comprehensive investigation into both the rhetorical structure (macro) and linguistic features (micro) of education RAs would make a significant contribution to our effort of assisting NNS (Non-Native Speakers) education researchers in crafting publishable RAs. It is precisely the pedagogical needs of guiding the rhetorically and linguistically disadvantaged NNS writers to communicating successfully that have motivated the ESP (English for Specific Purposes) approach to genre analysis.

Writing RAs in English has always been a daunting undertaking for NNS researchers, who encounter significantly more difficulties than native speakers of English (NS) when trying to produce texts that meet the requirements of the genre and the expectations of the readership (Burrough-Boenisch, 2003; Cho, 2004; J. Flowerdew, 1999). Specifically, insufficient knowledge of the generic rhetorical structure and the

conventions regarding lexical, syntactical and pragmatic choices to realize each rhetorical move has been identified as a major problem that places NNS writers at a distinct disadvantage when competing with NS writers for publication in the English language (Curry & Lillis, 2004; J. Flowerdew, 2001; Martin, 1985; Swales, 1990). The difficulties are sometimes so frustrating that NNS writers feel marginalized, or even excluded from the international scientific community.

1.2 Research Background

Over the past decades, Swales' pioneering work (1981, 1990, 2004) on the research article (RA), particularly his CARS (Create a Research Space) model for the introduction section, has not only stimulated enthusiasm for research on the RA and other academic and professional genres but also provided a valuable framework for later genre studies. Due to the central role it plays in disseminating research, the RA has received far more attention from genre analysts in the Swalesian tradition than any other academic genre. With the ultimate goal to use their research findings to inform syllabus design and materials development for students of English as a foreign or second language in different disciplines, these researchers have explored the rhetorical move structure of the specific sections of RAs in different disciplines, providing valuable insights into the intricate nature of the genre.

Two general characteristics can be noted of the literature. First, most studies have tended to focus on individual sections of the RA. Applying Swales' (1981, 1990,

2004) move analysis, numerous studies have examined the rhetorical structure of different sections of RAs. These include, for example, Swales and Najjar (1987), Samraj (2002), Hirano (2009), Loi (2010), Kanoksilapatham (2011) and Sheldon (2011) on the introduction section; Lim (2006) and Bruce (2008) on the methods section; Thompson (1993), Williams (1999), Bruce (2009) and Lim (2010) on the results section; and Hopkins and Dudley-Evans (1988) and Peacock (2002) on the discussion section. It is noteworthy that RA abstracts, too, have received much attention (Kanoksilapatham, 2009; Martín-Martín, 2003; Melander, Swales, & Fredrickson, 1997; Pho, 2009, 2010; Salager-Meyer, 1992; Stotesbury, 2003; B. Zhang, Thuc, & Pramoolsook, 2012). In contrast, the RA as an entire entity is under-investigated. To date, there have been only a limited number of investigations of RAs across all four sections (including Chang & Kuo, 2011; Kanoksilapatham, 2007, 2015; Nwogu, 1997; Pho, 2008a; Posteguillo, 1999; Tessuto, 2015). Clearly, further research into the rhetorical structure of the RA as a whole would be highly beneficial to NNS research writers.

The second characteristic of rhetorical structure analyses is a priority given to natural sciences over social sciences and humanities. Medicine (Nwogu, 1997; Salager-Meyer, 1992; I. A. Williams, 1999), biochemistry (Kanoksilapatham, 2007; D. K. Thompson, 1993), computer science (Posteguillo, 1999), and civil engineering (Kanoksilapatham, 2011), to name but a few, are among the many disciplines of natural

sciences or "hard sciences" investigated. In contrast, much less research literature exists with respect to "soft sciences." Among the few, are, for example, Brett (1994) on sociology; Holmes (1997) on history, political science and sociology; and Balocco (2000) on literature.

Interestingly, even within the area of social sciences and humanities relatively under-researched, a few studies have been conducted on different sections of education RAs. These studies cover a few specific fields of education research such as applied linguistics (Amirian, Kassaian, & Tavakoli, 2008; Lim, 2010; Pho, 2008a; Yang & Allison, 2003), educational psychology (Loi, 2010) and education technology (Pho, 2008a, 2010). However, as yet, no research has been conducted on RAs in education research as a general domain. Despite the interdisciplinary nature due to the integration of teaching and different academic disciplines taught such as science, health, languages, etc., different fields of education research have a common core to share—they are all educational in nature. Knowledge of the rhetorical structure that education RAs in specific areas are expected to share can have considerable pedagogical value for ESP and EAP (English for Academic Purposes).

While earlier research tended to analyze the rhetorical moves only, studies of linguistic features in RAs have been on the rise in recent years, either independent of or complementary to analyses of the rhetorical move structure. The driving force behind this intensified interest is undoubtedly provided by ever-maturing corpus techniques.

Move analysis explores the regularities of communicative purposes expressed by rhetorical moves. Corpus linguistics, too, depends on frequency counts to describe regularities of another aspect of genre—linguistic realizations of the communicative purposes. Studies of lexico-grammatical features based on specialized corpora of discipline-specific RAs have demonstrated the advantages of applying corpus techniques to genre analysis.

Corpus-based studies of lexico-grammatical features of RAs are characterized by two main focuses of attention: phraseologies and metadiscourse. Studies of phraseologies in RAs have investigated collocations and multiword sequences. Quantitative results have demonstrated that collocations and lexical bundles are not only central to academic discourse but, as in the case of the rhetorical structure, substantially variable across disciplines, between expert writers and novice writers, and among different sections of the RA (Cortes, 2004, 2008; Gledhill, 1995; Hyland, 2008a; G. C. Williams, 1998).

Based on the view of writing as a social engagement, research on metadiscourse has attempted to understand how authors organize their arguments and present themselves and their attitudes, and interact with their readers. Hyland's (1998b, 2005a) model of metadiscourse provides a comprehensive account of interactive and interactional metadiscourse devices. While interactive devices guide the reader through the text, interactional devices engage the reader in the propositional content by

indicating the author's attitude towards both the propositions and the reader. Again, in this strand of research on linguistic features, variations in the choice of metadiscourse devices have been shown to exist across disciplines (Hyland, 1998b) and among the different sections of RAs (Abdollahzadeh, 2011). In addition to inquiries into different metadiscourse devices, many studies have focused on specific lexico-grammatical items that realize those types of metadiscourse, for example, anticipatory *it* (Rodman, 1991), personal pronouns (Kuo, 1999), and the passive voice (Baratta, 2009).

While much of this previous work is characterized by a narrowness of focus on isolated, individual linguistic features, a shift towards investigating how sets of linguistic features are used in RAs is discernable in recent research into language use in RAs. For example, Swales (1990) examined how five linguistic features—*that* verb complement, present tense, past tense, passive voice, and hedging—combine in particular patterns in association with the rhetorical functions of the four standard sections of the RA. He observed a differential distribution of the lexico-grammatical features across the four sections.

A few studies have investigated systematic linguistic variation across the four sections of empirical research articles, assuming that rather than static, purpose is constantly responsive to immediate communicative needs (e.g., Biber & Finegan, 1994; Getkham, 2010). These studies took the advantage of corpus techniques, and applying Biber's multidimensional analysis framework (1986, 1988), identified a number of

functional dimensions underlying the co-occurrence patterns of linguistic features. However, few studies have gone further to analyze systematic linguistic variation at the move level. Indeed, no full-scale investigation into inter-move variation had been conducted until 2003 when Kanoksilapatham completed her doctoral dissertation on biochemistry RAs.

In conclusion, previous research on the RA has tremendously enhanced our understanding of the rhetorical structure of the RA and language choice in relation to the semantic units realizing the communicative purposes. However, certain substantial gaps are yet to be filled to benefit NNS academic writers and ESP/EAP practitioners as well. The major limitations of previous research range from an incomplete picture of the RA as a whole, to the narrowness of focus in analyzing linguistic features in the RA, to the lack of inter-move variation research, to the absence of research into education RAs as a genre in its own right representing the interface between teaching and a multitude of academic disciplines such as engineering education, business education, literacy education, applied linguistics and science education.

In view of the limitations, the study leading to this dissertation attempted to make a significant contribution to the knowledge base of the research genre, particularly education RAs, and in so doing, help linguistically and rhetorically disadvantaged educational writers in their competition for international publication.

1.3 Research Objectives and Questions

The overall purpose of this corpus-based investigation was to characterize the generic structure and language use of education RAs as a whole entity. This study integrated move analysis (Swales, 1981, 1990, 2004) and MD analysis (Biber, 1986, 1988) in investigating systematic linguistic variation among the rhetorical moves of education RAs, that is, how different patternings of linguistic features realize the different rhetorical moves across all the four IMRD sections of education RAs. Specifically, two objectives were attempted: 1) to identify the rhetorical move structure of education RAs; and 2) to map out linguistic co-occurrence patterns typically associated with the communicative purpose of each move in education RAs. The objectives translate into the following four research questions:

1) What are the rhetorical moves and constituent steps employed in each section of education RAs as represented by the ERC?

2) How are the moves and constituent steps typically sequenced in each section of education RAs as represented by the ERC?

3) What are the salient co-occurrence patterns of linguistic features in education RAs as represented by the ERC? and

4) What are the similarities and differences among the moves with respect to the use of the co-occurrence patterns of linguistic features?

1.4 Significance of the Study

This corpus-based study of the education RA genre can be beneficial in a number of ways. First, this study is expected to benefit both NNS education writers and ESP/EAP practitioners as well. By mapping out the move structure and clusters of lexico-grammatical features in the four sections of education RAs, this study provides insight into the scientific RA genre in general and in particular, education RAs hitherto largely neglected by genre analysts. NNS education writers will find the comprehensive descriptions of the move structure and linguistics features a useful aid in constructing effective RAs. In a similar vein, detailed and comprehensive delineations of the characteristics of this genre provide ESP/EAP practitioners with useful information for syllabus design and materials development.

Second, successful integration of qualitative move analysis and quantitative MD analysis in analyzing inter-move linguistic variation in education RAs provides further evidence of the advantages of the combination as initially experimented by Kanoksilapatham (2003). Since significant inter-move differences were found in terms of co-occurrences of linguistic features, move identification on the basis of communicative purpose is validated by the underlying dimensions of linguistic features that jointly realize the communicative function of each rhetorical move. The findings resultant from a combined use of move analysis and MD analysis can enhance validity in comparison with many previous studies that have relied solely on subjective manual analysis.

Finally, the corpus itself which served as primary data for this study has potential uses for future research and for ESP/EAP teaching. This study used a macroscopic approach to studying the overall rhetorical structure and constellations of linguistic features rather than individual ones. However, taking advantage of this corpus, a series of smaller-scale investigations can be undertaken to examine the behavior of individual linguistic features. Such microscopic investigations will complement macroscopic investigations by providing deeper insights into the language use of the education RA genre (Biber, 1988). With respect to learning and teaching purposes, concordances from this specialized corpus can prove to be a valuable resource for students in various education-related disciplines. For example, through observing concordances from this corpus, students can inductively infer the meanings and functions of individual words or lexical bundles in their authentic contexts. According to Johns (1994) who advocated data-driven learning, such discovery learning experiences boost learner motivation as well, particularly when they approach the corpus with specific questions.

1.5 Key Terms in the Study

The following are definitions of terms in alphabetical order that are either important or specific to the present study.

1) Education research: Education research is research carried out to improve educational policies, programs and practice (Martín-Martín, 2002; Mosteller et al.,

2004). Education research is characterized by its diversity as it embraces many traditions, paradigms, theoretical perspectives, methodological frameworks and academic disciplines. Broadly, the term covers both empirical research that relies on primary data and scholarly research that is entirely based on critical analysis of existing literature and theory. However, in a restricted sense, education research is "a disciplined attempt to address questions or solve problems through the collection and analysis of primary data for the purpose of description, explanation, generalization and prediction" (G. Anderson, 1998, p. 6). This study uses the term exclusively to refer to a systematic, empirical process of investigating a specific aspect of the educational system or education in general. Although the synonymous term "educational research" is used elsewhere, this study adheres to the term *education research*.

2) **Education Research Corpus (ERC)**: The Education Research Corpus (ERC) is a self-constructed specialized corpus of education RAs that served as primary data for the investigation of the target genre, the education RA. This genre-specific and goal-oriented corpus, with approximately a million running words (tokens), consists of 120 full-length empirical RAs systematically sampled from 12 peer-reviewed, high impact factor journals in education research.

3) **Inter-move variation**: Inter-move variation is the systematic linguistic variation among the different rhetorical moves. The term, especially coined for the present study, refers to the different ways particular sets of linguistic features (lexical,

grammatical, semantic or even pragmatic) are used in different rhetorical moves in education RAs. Inter-move variation was examined by comparing the mean dimension scores of the moves along the dimensions of linguistic variation. Knowledge of inter-move linguistic variation in education RAs can help NNS writers produce RAs that conform to the conventionalized ways of linguistic expression associated with the communicative purposes of the rhetorical moves.

4) Moves and steps: In genre analysis, the term *move* categorizes text segments conveying the same communicative purpose while the term *step* refers to a constituent part of a move (Swales, 1981). In this study, the minimum unit of move analysis was the sentence. Therefore, any sentence or group of sentences, irrespective of length, was considered an occurrence of a move as long as it had a dominant communicative purpose to realize. In cases where a sentence expressed more than one communicative purpose, only one move is identified by the most prominent purpose.

5) Multi-move sequences: In this study, a multi-move sequence in an RA section consists of a number of moves that co-occur saliently more frequently than would be expected by chance. Multi-move sequences are also known as "move bundles" (Chang & Kuo, 2011), "move collocations" and "move clusters" (Feng, 2006). This concept is useful in that frequency information of sequences of moves can be a basis for determining the relative positions of the moves in an RA section and thereby deriving the section's prototypical move structure. Technically, multi-move sequences were

extracted from the corpus using the n-gram function of the concordance program AntConc (Anthony, 2014).

6) Multidimensional analysis: Multidimensional (MD) analysis, developed by Biber (1986, 1988), is a corpus-based quantitative approach to the analysis of linguistic variation among registers or genres. Two assumptions underlie this analytical framework. First, certain linguistic features tend to occur together because they share similar or closely-related functions. Second, language varieties such as different languages dialects, speech, writing, genres, and registers are multidimensional in terms of communicative purpose, and as such, they should be described comprehensively with respect to how co-occurring linguistic features communicate the purposes. The core of this analysis is using factor analysis to identify patterns of co-occurrence of a broad range of linguistic features in text thus revealing the major dimensions on which texts vary.

7) Part of Speech tagging: Part of speech tagging, also known as grammatical tagging or morpho-syntactic annotation, assigns part of speech tags to each word in a corpus (Hunston, 2002). Unlike an untagged, raw corpus which, to a large extent, limits searches to literal words, a tagged corpus allows inquiries into the behavior of lexical classes, grammatical categories and syntactic constructions. The ERC on which this study was tagged with the Stanford POS Tagger (Toutanova et al., 2015).

1.6 Summary

This chapter identified the pedagogical needs of helping NNS education researchers write RAs as the general research problem. Then in outlining the ESP approach to genre analysis and most relevant past literature, some major gaps in existing research literature were exposed. They include incomplete generic and linguistic accounts of the RA genre and a neglect of education RAs as an entire entity in past research. Thus, the general research problem was gradually narrowed down to a specific need to investigate the move structure and inter-move linguistic variation of education RAs. Subsequently, the objectives, design and significance of the research were outlined. In addition, definitions were provided for the important terms used in this study.

Chapter 2 will present a critical review of theories and empirical studies relevant to the present study.
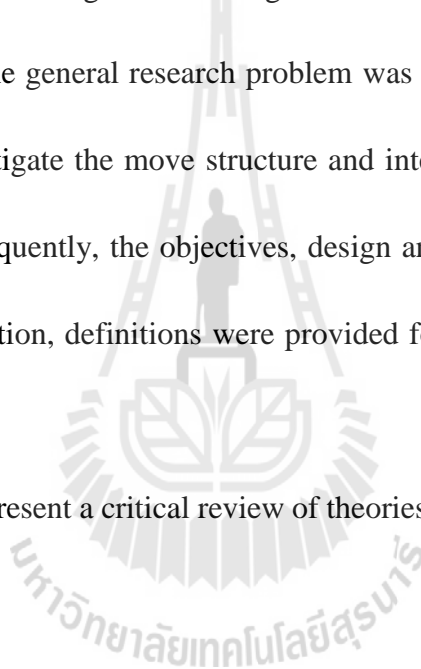
CHAPTER 2

LITERATURE REVIEW

This chapter reviews previous literature, both theoretical and empirical, that is relevant to the study. Section 2.1 presents an overview of genre theory and the different approaches to genre, but the emphasis is on the ESP approach as it provides the analytical framework for the move analysis part of the study. Section 2.2 reviews previous genre research studies of RAs within the ESP tradition. Section 2.3 concerns corpus linguistics theory, with a focus on the interface between corpus linguistics and genre analysis. Subsequently, corpus studies of linguistic features in RAs are presented in Section 2.4. Section 2.5 has two concerns: MD analysis which was the framework for the linguistic analysis part of this study, and Kanoksilapatham's (2007) MD study of inter-move variation. Section 2.6 is an overall critique of previous empirical research on RAs. The chapter concludes with Section 2.7 that establishes the connection of the literature with the present research.

2.1 Genre Theory

Genre has a long-established tradition in literature, but only in recent years it has become such a prevalent notion in language teaching and learning that it represents an important "paradigm shift" (A. M. Johns, 2002, p. 3). Genre analysis holds so much

appeal for language teachers because it is truly applied in that, by exploiting the conventional aspects of language use, it answers the question why members of a discourse community construct their texts the way they do. What enables genre analysis to answer this question is the attention given to the broad context in which discourses are produced. By providing socio-cultural, institutional and organizational explanations, genre analysis has successfully transcended mere surface-level grammatical description and has arrived at "significant form-function correlations" (Bhatia, 1993, p. 11). Such form-function descriptions can inform language teaching, especially ESP and EAP.

Over the years, three general, overlapping approaches to genre analysis have emerged, including the ESP approach, the Sydney School and the New Rhetoric (J. Flowerdew & Wan, 2010; Hyland, 2002b; Hyon, 1996). These approaches, united by a common goal to describe and explain regularities of purpose, form and context, can be distinguished by the weight they attach to text or context.

The following subsections discuss these approaches, with their common literary roots introduced first in passing. However, the ESP approach, which provides the analytical framework for the move analysis part of this study, is given more space, whereas the other two receive a less detailed treatment.

2.1.1 The Literary Tradition

Although it is as old as thought itself (Herrington & Moran, 2005), the concept of genre can be traced back at least to Aristotle, who defined and described epic, lyrical, and dramatic poetry by identifying their constituent parts (Bawarshi & Reiff, 2010; Santini, Mehler, & Sharoff, 2010). For example, Aristotle stated that a tragedy should have six conventionalized constituents, namely, plot, character, diction, thought, spectacle, and song. Essentially, the concept served to classify literary texts by their formal properties. In the 18th century, English classical commentators inherited this concept and started to use the French word "genre" meaning "kind" or "class" to classify different types of literary texts and later other forms of artistic production. For a long time genre was regarded as discrete and stable despite the fact that writers often mixed the genres, as illustrated by Shakespeare's *Macbeth* and *Hamlet* that combined comedy and tragedy. In the 19th century the concept of genre gradually fell from favor with the literary critics, largely due to the emergence of new genres such as the novel and to their preference for originality over conventionality (Corbett, 2006). As a result, in literary studies and in other aesthetic fields, such as art and film criticism, a work that strictly adhered to the prescribed form was considered having low aesthetic value and was often contemptuously called a "genre piece."

However, even in the late 19th century and early 20th century, genre was kept alive in the composition classroom, mostly in the form of "modes of discourse":

exposition, persuasion, description, and narration (Herrington & Moran, 2005). Rather than recurring social situations, the rationale behind these "modes" was the so-called four mental faculties, namely, understanding, imagination, emotion, and will. Because of these faculties, the four purposes of discourse were possible: to inform, to please, to arouse emotion, or to influence action. For a long time, the prescriptive, a-contextual "modes" pedagogy dominated the composition classroom "through complete control of textbooks" (Connors, 1981, p. 449).

2.1.2 The ESP Approach

Although genre theory is profoundly rooted in a tradition obsessed with linguistic form, the revival of genre since the 1980s is marked by a turn towards functionalism as the rationale behind genre. Of all the three contemporary approaches that take into consideration both text and context, the ESP approach is perhaps the most influential.

It was Swales (1981, 1990, 2004) who formulated this approach. In his famous book on genre analysis, Swales (1990) provided this classic definition of genre：

> *A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style. Communicative purpose is both a privileged criterion and one that operates to keep the scope of a genre as here conceived narrowly focused on comparable rhetorical action. In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience. (p.58)*

Here the centrality of the communicative purposes of the discourse community is highlighted. A genre is primarily a category of discourses identified by shared communicative purposes that are recognized by the discourse community. Due to their common purposes, instances of a genre have a conventionalized rhetorical structure and ways of linguistic expression.

Swales' work centered on genres in academic settings such as RAs, grant proposals, dissertations, and oral presentations. In particular, his 3-move CARS model (Swales, 1981, 1990, 2004) for RA Introductions has become a standard model for move analysis.

Bhatia (1993) emphasized that it is the shared communicative purposes that distinguishes genres and make them mutually understandable, and that genre frequently has a highly conventionalized internal structure. To be accepted as a full member of the discourse community, one must have adequate generic knowledge or knowledge of the conventions of discourse in the relevant discourse community. Findings from analyses of the conventionalized rhetorical structures of different academic and professional genres, therefore, can inform ESP/EAP syllabus design and materials development.

Since genre is purpose-oriented, genre analysis focuses on the rhetorical moves or strategies that realize the overall communicative purpose. The sequence of moves and steps may vary between different instances of a genre. Some moves may be

optional; some may occur in a different order; some may be embedded in others; and still some may be recycled. However, an invariant prototypical move structure or the overall organizational pattern of the moves and steps can always be identified. Meanwhile, in correspondence to each of the moves and steps, typical conventionalized ways of expression are used to realize the specific purpose of the move or step. Genre analysis, sometimes called move analysis, therefore, involves identifying the moves or strategies of a genre, determining the allowable order of the moves based on frequencies, and analyzing the key linguistic features that realize the moves. The next step is to explain why these features are chosen by expert users of the genre to achieve their communicative purposes (Bhatia, 1993).

Categorizing texts in terms of communicative purpose is believed to have the advantage of turning teachers' and students' attention away from mere surface structures of text to the socially situated use of texts with specific intentions (Hüttner, Smit, & Mehlmauer-Larcher, 2009), but until recently, many ESP researchers paid particular attention to the formal characteristics of genres while focusing less on the specialized functions of texts and their surrounding social contexts (Hyon, 1996). A better balance of form and context has been achieved in more recent genre studies. For example, Bhatia (2002) emphasized the importance of intertextuality, the notion that every text is "bound in relations to other texts which are somehow present in it and from which it draws its meaning, value, and function" (Venuti, 2009, p. 157).

Bhatia (1993) extended Swales' scope to include professional settings, sub-genres, and mixed genres. For example, within the RA there are such embedded genres as the survey article, the review article, and the state-of-the-art article. Bhatia (1993) compared the move structures of job application letters and sales letters, and found that the two share the same communicative purpose thus belonging to the same promotional genre.

Adimittedly, move analysis has some drawbacks. Traditonally, it has focused on the rhetorical organization of texts and exclusive attention has been given to linguistic structures beyond the sentence level. Furthermore, the qualitative nature inevitably gives a subjective color to move analysis. Move analysis is basically qualitative because usually a small number of texts from a single genre are examined in detail (Biber, Connor, & Upton, 2007). It is qualitative also because moves are often intuitively identified by examining the communicative purpose of text segments.

2.1.3 The Sydney School

The Sydney School, which originated at the University of Sydney, employs a methodology derived from Halliday's Systemic Functional Linguistics (SFL) (Halliday, 1994; Halliday & Hasan, 1989). Interest in genre arose out of growing dissatisfaction with register analysis that could not provide an adequate explanation for the contextual aspects of text. The realization that a text might be shaped and so categorized by socio-cultural context led to a resurgence of interest in a text's genre. Genre is seen as

"a staged, goal oriented social process" (Martin, 1992, p. 505). SFL research has given much weight to the social purposes of genres and to revealing the rhetorical structures that serve those purposes. Therefore, texts that function to serve the same social purpose will often share the same rhetorical structure, and thus constitute a genre.

The relationship of text and context is central to this approach. Each context has the possibility for a range of possible texts called Generic Structure Potential (Hasan, 1989). Register and genre are the two levels at which texts are connected to particular contexts. When creating a text, people first make choices in register. Three elements are involved at this level: field, tenor and mode. Field refers to the social activity in which one is involved and what the text is about. Tenor is the relationships of the participants in the interaction. Mode is the role of language (e.g., wholly written, written and spoken, illustrations). At the genre level, linguistic choices are influenced by the writer's social purpose in using language. Whereas register is associated with broad fields of activity, genre is a more concrete expression of field, tenor, and mode, involving conventions for organizing messages so that readers can recognize the purposes (Hyland, 2004).

Unlike the ESP approach which groups texts as genres, for example, RAs and recommendation letters, by regularly occurring activities, the Sydney School classifies texts into genres by internal linguistic criteria. With its origin in a linguistic framework, this approach tends to characterize genres in terms of broad rhetorical patterns such as narrative, recounts, arguments, and expositions.

2.1.4 The New Rhetoric

The New Rhetoric is concerned with composition studies and professional writing in an English L1 context. Clearly, the ESP and Sydney School approaches are fundamentally linguistic in orientation as they both emphasize communicative purpose, schematic structure and form-function connection. The New Rhetoric, however, while acknowledging the regularities and conventions within genres, places greater emphasis on the dynamic quality of genres as they are more "flexible, plastic, and free" than assumed (Bakhtin, as cited in Hyland, 2004). Factors that give rise to variation within genres are thus to be accounted for. Within this school, in addition to describing rhetorical patterns and linguistic forms, genre analysts approach genre analysis ethnographically and focus on the activities, beliefs, values and patterns or behavior of the discourse community (J. Flowerdew & Wan, 2010). Consequently, participant observation, interviews, and descriptions of physical settings, as well as analysis of texts, are usually involved in genre studies (Hyland, 2004).

In conclusion, the three approaches to genre provide useful answers from different angles to the questions of how genres are constructed and why they are constructed the way the are. The ESP approach attaches relatively more importance to the internal rhetorical structure of genre as motivated by communicative purpose. In addition, its move-step analytical model has been proven powerful in characterizing the rhetorical structures of many academic and professional genres.

In view of the purpose of the present study, move analysis of the ESP approach was adopted as an analytical framework to analyze the rhetorical structure of the education RA genre. The Sydney School, which categorizes genre in more linguistically functional terms, indicates that analyzing inter-move linguistic variation of education RAs may be a worthwhile effort. In the conduct of move analysis, inspiration was also drawn from the New Rhetoric. Although it was impossible to carry out participant observations and interviews with the actual authors of the texts in the corpus, examining the context beyond the text (e.g., submission guidelines, introductions to journals, and author bios) and enlisting the assistance of another education researcher to code some of the education RAs contributed to the validity of results.

2.2 Empirical Move Analyses of RAs

Swales' CARS model (1981, 1990, 2004) for RA introductions has stimulated numerous move-based investigations on different academic and professional genres. However, the RA is by far the most extensively investigated. Swales himself has refined his analysis of RA Introductions (1990, 2004) and other sections of the RA have been subjected to move analysis. While early studies tended to analyze the move structure alone, more recent research has adopted a corpus approach to investigate language use in RAs to complement descriptions of the overall organizational structure.

This section reviews studies of the generic move structure of RAs whereas studies of linguistic features are examined in the section immediately following this. The rationale behind this separate treatment is that move analyses and analyses of linguistic features represent two distinct levels of textual analysis, with the former focusing on features beyond the sentence and the latter within the sentence.

While most previous move-based studies of the RA have analyzed the rhetorical move structure of independent sections of RAs, only a limited number have attempted to examine the RA as a whole entity. Therefore, it is reasonable to classify previous studies into categories corresponding to each of the IMRD sections of the RA. Studies of the rhetorical structure across all four sections are reviewed in a separate subsection.

2.2.1 The Introduction Section

The introduction section welcomes the reader to the article by providing brief but convincing answers to questions concerning the importance of the research problem, the relationship between the study and previous work in the area, the primary and secondary hypotheses and objectives and their relevance to theory, the research design in relation to the research questions, and the practical and theoretical implications of the investigation. By summarizing the relevant arguments and the previous research, an introduction addresses these questions and enables the reader to understand what was done and why (American Psychological Association, 2010).

The introduction section is the birthplace of a whole enterprise of genre research of the ESP tradition. Back in 1981, Swales examined the introduction sections of 48

articles in the natural and social sciences, and empirically derived a 4-move structure to account for the rhetorical moves and steps that a research writer typically employs in realizing his or her communicative purpose. He metaphorically called this 4-move structure the Create-a-Research-Space model to liken the fierce competition for academic promotion to the competition for a living space in the ecosystem. Later he fine-tuned the model twice (Swales, 1990, 2004) to accommodate new findings by himself and by other researchers (e.g., Crookes, 1986; Jacoby, 1986; Kwan, 1996; Lopez, 1982). The most significant modification was the merge of the first two moves in the original model: *Establishing a territory* and *Summarizing previous research*. However, despite the modifications, subsequent research on the rhetorical move structure of RA introductions has typically favored the 1990 model. A major reason why few studies have used the 2004 model as a framework is that, as Del Saz-Rubio (2011) explained, it basically describes the underlying structure of the 1990 model, and outlines possible additional steps within the moves without radically modifying them.

Figure 2.1 presents Swales' 1990 3-move CARS model that has been successfully applied to many subsequent studies. In the opening *Establishing a territory* move, the writer situates his or her own work in a broad field of research by following one of these steps: pointing out the importance of the general subject, making general statements about the subject, and reviewing items of previous research. Then the writer makes the second move *Establishing a niche* to indicate the specific area of the broader subject that the article will deal with. This is accomplished by making an

assertion contrary to expectations, or exposing a deficiency of previous research, or raising a question about existing research, or suggesting that the present study is following in the footsteps of previous researchers, or the writer may take a combination of several of those steps. In the final move *Occupying the niche*, the writer sketches out exactly what the present study will accomplish in relation to the second move, and gives the reader a sense of how the article will proceed. The writer takes these steps in order: stating the purposes of the article or outlining the main features of the present research, summarizing the major findings/results of the study, and signaling the organizational structure of the article.

Move 1: Establish a territory
    Step 1   Claiming centrality, or and/or
    Step 2   Making topic generalization(s), and/or
    Step 3   Reviewing items of previous research
Move 2: Establishing a niche
    Step 1A Counter-claiming, or
    Step 1B Indicating a gap, or
    Step 1C Question-raising, or
    Step 1D Continuing a tradition
Move 3: Occupying the niche
    Step 1A Outlining purposes, or
    Step 1B Announcing the present research
    Step 2   Announcing principle findings
    Step 3   Indicating RA structure

Figure 2.1 Swales' CARS Model for RA Introductions

Swales' two-level CARS model of moves and steps has provided a powerful analytical framework for later move analyses, resulting in voluminous literature on various academic and professional genres. Yet, RA introductions have received much more attention than any other genre or sub-genre.

Swales (1990) suggested that the CARS model could account for the structural organization of RA introductions irrespective of discipline. However, results of subsequent research have indicated considerable disciplinary and cross-linguistic variations. In fact, disciplinary and cross-linguistic variations have been two recurrent themes in subsequent research on RA introductions ever since Swales' CARS model was formulated.

Variation of the rhetorical move structure of RA introductions is reflected across different academic disciplines. Cross-disciplinary variation studies have shown that, though in general conformity with the CARS model, RA introductions of different disciplines deviate in varying degrees from the model (e.g., Anthony, 1999; Crookes, 1986; Del Saz-Rubio, 2011; Kanoksilapatham, 2007; Posteguillo, 1999; Samraj, 2002; Swales & Najjar, 1987). Most such studies have identified additional features or phenomena not accounted for by the CARS model.

Crookes' (1986) found that in contrast to hard sciences and biology/medical sciences, social sciences RA introductions are lengthy and occasionally contain topic-specific subheadings. In terms of move structure, introductions in both "hard" sciences and biology/medical sciences have a closer affinity with the CARS model than social sciences that tend to recycle the moves.

In another study, Anthony (1999) observed important features in 12 software engineering RA introductions that did not fit into the CARS model, such as the

presence of many definitions of terms and exemplifications of difficult concepts in Move 1, and an evaluation of the research in Move 3. Similarly, in her study of wildlife behavior and conservation biology RA introductions, Samraj (2002) identified a background move in wildlife behavior RA introductions concerning features of the species under investigation. Moreover, she found that the review of literature step, rather than being limited to Move 1, could occur in all three moves in both disciplines. Interestingly, Del Saz-Rubio (2011) linked particular step combinational patterns to different communicative purposes.

Disciplinary variation has also been found among closely related sub-disciplines within a discipline. However, studies in this respect are relatively scanty. One example is Ozturk's (2007) comparative study of 20 RA introductions in L2 acquisition research and L2 writing research. This study revealed marked differences between the structural organization of the two sub-disciplines of applied linguistics. Her results indicated that while the move structure M1-M2-M3 described in the CARS model was predominant in L2 acquisition research, the M1-M2-M1-M3 and the M1-M3 patterns were almost equally frequent in L2 writing research. Ozturk attributed this inter-disciplinary variation to the fact that L2 acquisition research is an established field whereas L2 writing research is still emerging.

From a cross-linguistic perspective, research on variation between RA introductions in different languages has identified some culture-specific characteristics.

For example, Fakhri's (2004) study of Arabic RA introductions in the humanities and social sciences exhibited a substantial divergence from the CARS model. He attributed this divergence to the authors' academic background and the lack of a competitive research environment. Another example is Hirano's (2009) analysis of English and Brazilian Portuguese applied linguistics RA introductions. Hirano found that her English texts generally conformed with the CARS model whereas the Brazilian Portuguese texts lacked a gap statement. Two more recent studies (Loi, 2010 on educational psychology; Y. Zhang & Hu, 2010 on medicine) compared introductions in English and Chinese. Both found that although the global rhetorical organization of the Chinese texts generally resembled that of the English introductions, Chinese writers tended to use Move 2 *Establishing a niche* much less frequently than their English counterparts. In addition, Loi observed a total absence of certain steps in the Chinese introductions, for example, the *Presenting the theoretical basis* step in Move 1 and the *Counter-claiming* step in Move 2.

While most cross-linguistic variation studies have compared two different languages, Sheldon's (2011) investigation represents an interesting three-way comparison. She investigated 54 applied linguistics RA introductions in English and Spanish as well as English introductions authored by Spanish-background speakers. The English L1 texts were found to conform with the CARS schema, but the Spanish L1 texts exhibited some distinct culture-specific features. The English L2 introductions,

however, had a greater affinity with the CARS model as they were intended for an international readership.

In sum, much of previous research on RA introductions is characterized by a comparative perspective. Since Swales' CARS model (1981, 1990, 2004) was first proposed, nearly all studies of the introduction section have compared introductions between different disciplines, sub-disciplines, and languages and have applied the CARS model as a baseline for comparison. The availability of a model purported to account for all introductions regardless of discipline is an important reason for the abundance of studies on this section.

2.2.2 The Methods Section

The overall communicative purpose of the methods section is to present a detailed description of the method employed in the study so that the reader can evaluate the appropriateness of the method and the reliability and validity of the results thus produced (American Psychological Association, 2010). The diversity and complexity of different disciplines give rise to different types of research employing different methodologies which in turn result in methods sections whose rhetorical move structure is difficult to characterize. This is one reason why studies of the rhetorical structure of this section are relatively few (e.g., Kanoksilapatham, 2007; Weissberg & Buker, 1990). In addition, a lack of discipline-specific knowledge may have, to a great extent, discouraged attempts to analyze this section.

Nevertheless, a few discipline-specific studies have been conducted on the methods section. Wood (1982) identified a 3-move structure for chemistry RA methods sections: 1) Describing the sample, 2) Describing an apparatus, and 3) Describing experimental procedures. He also noticed that Move 2 and Move 3 were optional in that they might not be used if a standard apparatus or established procedures were used in the experiment.

Like Wood's (1994) account of methods sections in chemistry, both Brett's model for sociology and Nwogu's (1997) for medical research also contained three similar moves. According to them, sociology and medical RA methods sections start with the same move *Describing data collection procedures* and end with the same move *Describing data analysis procedures*, but in the second move, sociology methods sections focus on explaining the concepts and variables of the research work whereas medical methods sections describe experimental procedures. The difference can be explained by the fact that sociological studies frequently depend on non-experimental methods like surveys and case studies to establish correlations between various variables while medical research typically involves experiments to establish causal relationships between variables.

Lim (2006) found that most methods sections from a corpus of management RAs contained three major moves: 1) *Describing data collection procedures*, 2) *Delineating procedures for measuring variables*, and 3) *Elucidating data analysis procedures*. Each

of these moves was realized by several steps. His model resembles previous accounts of

the methods section, but one particular step in Move 3 is noteworthy. After recounting

and justifying data analysis procedures, the researcher may take an optional step

Previewing results to report data that are treated as preliminary results to be further

interpreted to produce specific findings. Methodologically, Lim's study is also

noteworthy in that he, taking an ethnographic perspective, held informal interviews

with expert informants to validate his findings.

Though scanty, previous studies on the methods section, have revealed some

disciplinary variation of the rhetorical structure of the section. While a uniform model

comparable with the CARS model to characterize the methods sections in all

disciplines seems unlikely, NNS writers and ESP/EAP practitioners would welcome

detailed information on the conventional ways of constructing the methods sections of

specific disciplines. Such information can only be obtained by analyzing methods

sections specific to their disciplines or the methodology they use.

2.2.3 The Results Section

The results section functions to report the results of a study in an accurate,

unbiased, and complete manner. This section is crucial because it "drives the paper and

therefore should be written first" (Cargill & O'Connor, 2006, p. 210). The "Publication

Manual of the American Psychological Association" (American Psychological

Association, 2010) advises the researcher to: 1) summarize the data and the analysis of

data, 2) report the data in sufficient detail to justify conclusions, and 3) present all relevant results including the unexpected or undesirable.

Of the previous studies on or including the results section from a move analysis perspective, Brett's study (1994) is probably the most important. In 20 sociology RA results sections, Brett identified 13 rhetorical moves under three communicative categories: Metatextual, Presentation and Comment. The Metatextual category embraced statements that referred the reader to the data or to other sections in the article; the Presentation category included objective and impersonal statements to report, present, or highlight the results or the ways in which they were obtained; the Comment category covered statements whereby the researcher offers his or her own interpretation of or comment on the results already presented. Brett further observed that the occurrence of the three organizational categories was cyclical, the most frequent pattern being *Pointer* (Metatextual) followed by *Statement of finding* (Presentation), and *Substantiation of the finding* (Presentation).

Posteguillo (1999), taking Brett's (1994) model as a reference, examined 22 results sections in computer science. His results indicated that computer science RA results sections consisted of 10 moves under the three communicative categories of Metatextual, Presentation, and Comment moves. The metatextual moves were Move 1 *Pointer*, and Move 2 *Structure of section*. The presentation moves included Move 3 *Procedural*, Move 4 *Hypothesis restated*, and Move 5 *Statement of data*. The comment

category had Move 6 *Comparison of finding with literature*, Move 7 *Evaluation*, Move 8 *Further research suggested*, Move 9 *Implications*, and Move 10 *Summarizing*. Furthermore, move recycling was found to be common in this section. The most charateristic cyclical patterns consisted of these combinations of moves: Procedural-Pointer-Statement of data or Procedural-Pointer-Evaluation of data.

Another study that drew on Brett (1994) is Williams (1999), who applied a modified version of Brett's model in his analysis of medical RAs. Ten moves were identified, including 1) *Pointer*, 2) *Structure of section*, 3) *Procedural*, 4) *Statement of finding/result*, 5) *Substantiating finding*, 6) *Non-validation of finding*, 7) *Explanation of finding*, 8) *Comparison of findings with literature*, 9) *Evaluation of finding re hypotheses*, and 10) *Interpretation of finding*. Brett's model for sociology results sections was found to be basically valid for interdisciplinary genre analysis. However, the cyclical patterning of moves identified in Brett's sociology RA results sections was less common in the medical research RA results sections, suggesting that text organization is constrained by both type of study and subject matter.

Brett (1994), Posteguillo (1999), and Williams (1999) have shown that the results section is not only where new findings are highlighted but also where they may be interpreted and commented, often in a cyclical fashion. The presence of commentary moves in the results section highlights the complex relationship between the major RA sections. This complexity was noticed by Swales (1990) who pointed out that results

and discussion sections are sometimes combined and that sometimes sections labeled

Conclusions, Implications or Applications are added.

This intriguing complexity prompted Yang and Allison (2003) to examine how

neighboring sections have a bearing on the rhetorical structure of the results section,

rather than treating the section as an independent entity as in Brett (1994), Posteguillo

(1999), and Williams (1999) and many other studies. Yang and Allison analyzed the

mutual influence of results, discussion and conclusion sections. Significantly, their

findings, based on a sample of 20 applied linguistics articles, indicated that, due to the

overlapping of rhetorical functions of the different sections, the structure of empirical

RAs in applied linguistics tended to be flexible as the article proceeded towards the end.

However, as they pointed out, the sections were still distinct in terms of their primary

communicative purposes and this generally motivated the use of different section

headings. With respect to the results section, they observed that results sections in

applied linguistics generally had a highly cyclical structure, and they not only reported

results but also briefly commented on them. This finding is in keeping with those

studies in other disciplines such as Brett (1994), Posteguillo (1999) and Williams

(1999). Yang and Allison proposed a 6-move structure for results sections: 1)

Preparatory information, 2) *Presenting results*, 3) *Summarizing results*, 4)

Commenting on results, 5) *Evaluating the study*, and 6) *Deductions from the research*.

Among them, the first three are dominant moves.

The studies reviewed above and others not reviewed (e.g., Kanoksilapatham, 2007; D. K. Thompson, 1993) provide much insight into the rhetorical structure of the RA results section. While there are marked similarities in the models proposed, there also appears to be disciplinary variation. However, most disciplines share the two conventional moves of *Presenting results* and *Commenting on results* which may occur in a linear manner but typically in a cyclical way.

2.2.4 The Discussion Section

The discussion section discusses the results presented in the results section by evaluating them and interpreting their implications, especially with respect to the research questions (American Psychological Association, 2010). Since it is subject to the most critical reading by research consumers, this section is thought to be the most important section in the RA, and hence the most troublesome to write (Berkenkotter & Huckin, 1995).

Because of its importance and the challenge faced by NNS researchers in writing it, previous literature is almost comparable in volume to that on the introduction section. Like in studies of the other RA sections, the focus of research, again, has been on disciplinary and cross-linguistic variation regarding the rhetorical structure of this section.

Belanger's investigation (as cited in Swales, 1990) can be considered a pioneering study on this section. His analysis of neuroscience discussion sections

established a close correlation between the rhetorical structure of this section and the number of research questions the study attempted to answer. His model for this section has 5 moves: 1) *General introduction*, 2) *Summarizing results and stating conclusions with references to previous research*, 3) *What results suggest with references to previous research and/or to the current work*, 4) *Further questions with possible explanations or with references*, 5) *General conclusion*. Among them, Moves 2, 3, and 4 can be recycled as many times as there are research questions.

Hopkins and Dudley-Evans (1988) examined an unspecified number of MSc dissertations in biology and conference papers on irrigation and drainage. The result was an 11-move model: 1) *Background information*, 2) *Statement of result*, 3) *(Un)expected outcome*, 4) *Reference to previous research*, 5) *Explanation of unexpected results*, 6) *Exemplification*, 7) *Deduction*, 8) *Hypothesis*, 9) *Reference to previous research*, 10) *Recommendation* and 11) *Justification*. Besides, the researchers identified in the conference papers a typical pattern of three cycles. They explained that "emphasis in these cycles is very much on the interpretation of results, and the discussion sections of articles and dissertations appear to be judged less on the actual results presented than on the way the writer relates them to previous work in the field" (Hopkins & Dudley-Evans, 1988, p. 119). This observation lends strong support to Belanger's (as cited in Swales, 1990) claim about move recycling in the discussion section.

Swales (1990) noted that Hopkins and Dudley-Evans' (1988) 11 moves were almost identical with those 11 moves in Peng's (1987) model for chemical engineering articles. Therefore, he proposed a distilled model consisting of the following 8 moves: 1) *Background information*, 2) *Statement of results*, 3) (Un)expected outcome, 4) Reference to previous literature, 5) *Explanation*, 6) *Exemplification*, 7) *Deduction and hypothesis*, and 8) *Recommendation*.

Two studies by Basturkmen (2009, 2012) focusing on the key move of *Commenting on Results* are particularly revealing. In her 2009 study of applied linguistics RA discussion sections, Basturkmen identified four moves: 1) *Background information*, 2) *Summarizing results*, 3) *Reporting a result*, and 4) *Commenting on the result*. In her 2012 study of dentistry RA discussion sections, she identified two additional moves after the commentary move: *Evaluating the study* and *Implications for further research and/or clinical practice or policy*. Furthermore, in both studies, she identified a result-comment pattern repeated as many results as the writer wished to discuss. As for steps of the commentary move, she found in both studies that the writers elected from three possible steps: *Explaining a result*, *Comparing a result with a result reported in the literature*, and *Evaluating a result*.

Other researchers, for example, Berkenkotter and Huckin (1995), have suggested that the many moves in the discussion section can be actually ordered into a set of higher level units that mirror the moves in the CARS model (Swales, 1981, 1990, 2004)

for the introduction section. In other words, the moves are essentially the same as those in the introduction, but the order is reversed: 1) O*ccupying a niche (Statement of principal findings)*, 2) *(Re)establishing the field (Show how results respond to the larger issue stated in the introduction)*, and 3) *Establishing additional territory (Implications of the study and sometimes plans for future work)*. According to them, the function of this structure is to situate new research findings within the body of knowledge previously accepted by the relevant community.

Meanwhile, a few studies have explored cross-linguistic variation in discussions. Williams (2005) compared medical science RA discussions in Spanish and English. To his data, he applied this move-based framework: 1) *Statements providing background information*, 2) *Expressing results*, and 3) *Comparing current and previous findings, or Making knowledge claims*. The results indicated that Spanish writers tended to use a progressive pattern whereas English writers preferred a retrogressive pattern.

Another study across languages is Amirian, et al. (2008), which examined three corpora of applied linguistics RAs: English L1 RAs, Persian L1 RAs, and English L2 RAs by Persian researchers. Using Hopkins and Dudley-Evans' framework (1988) in their analysis, they found considerable differences across the three corpora despite their similarities in move structure. Interestingly, three moves unique to the Persian L1 and English L2 texts were observed, including *Hedging statement*, *Reference to previously mentioned statement*, and *Expressing wish for further research*. This finding suggested

cultural influence on the rhetorical structure. Moreover, based on their English L1 data, they proposed a model of 10 micro-moves ordered into 3 macro-moves. The macro-move of Introduction contained three micro-moves: 1) *Presenting background*, 2) *Reference to previous research* and 3) *Statement of aims*; the Body also had three: 1) *Findings*, 2) *Explanation*, and 3) *Reference to previous research*; the Conclusion had four: 1) *Restatement of findings*, 2) *Reference to previous research*, 3) *Limitations of study*, and 4) *Recommendations for further research*. The researchers further pointed out that some moves in this model could be combined in different ways, often in cyclical patterns, according to the writer's intentions.

In a recent study, Amnuai and Wannaruck (2013b) compared applied linguistics RA discussion sections by international and Thai writers. They found that both groups of writers employed the same 7-move scheme in Yang and Allison' model (2003). However, they observed that Thai writers employed the *Deduction from research* move significantly more frequently than international writers because they wished to attract attention to the value of their research, particularly its practical implications. As for which move to open the discussion section, they found that international writers favored the background information move while Thai writers the results reporting move.

Previous research on the RA discussion section has produced multiple rhetorical structures for the section, but never a uniform model applicable to all disciplines and all

types of research. Given the complexity of the discussion section due to disciplinary and methodological differences, more research from a cross-disciplinary perspective will certainly be profitable.

2.2.5 Entire Research Articles

Recent years have witnessed increased attention to all-four section RAs as researchers realize the importance of understanding the rhetorical structure of the RA in its entirety. The rhetorical structures of a range of disciplines have been investigated from different perspectives. Some focused on a single discipline; some had a cross-disciplinary or cross-linguistic perspective; one even had a diachronic perspective. However, as with analyses of discrete RA sections, investigations into whole RAs exhibit the same strong bias towards "hard" sciences. Studies on "hard" disciplines far outnumber those in "soft" disciplines.

Among the "hard" disciplines investigated were medicine, computer science, engineering, chemistry, and agriculture. Medical RAs were the first and most investigated. The three studies on medical RAs (ElMalik & Nesi, 2008; Li & Ge, 2009; Nwogu, 1997) unanimously revealed an 11-move rhetorical organization. In fact, the latter two adopted the model developed in the first study. Yet, differences were found in move status. For example, in Nwogu's (1997), *Presenting background information* and *Describing data analysis procedures* were optional moves. Li and Ge's (2009) comparative analysis of articles from two different historical periods revealed that the

two moves had changed from optional to conventional whereas *Highlighting overall research outcomes* had gone the other way round. The differences were attributed to the evolution of medical science and changes in attitude on the part of medical RA writers.

The two studies on computer science RAs (Chang & Kuo, 2011; Posteguillo, 1999) revealed that computer science RAs often did not fit into the conventional IMRD structure. Posteguillo found that instead of having the conventional free-standing "Methods" section, computer science articles often featured such sections as "Preliminaries," "Algorithms," and "Analysis of a Problem" between the introduction and results sections. Similarly, Chang and Kuo observed that in computer science RAs, the methods, results and discussion sections were often mixed between the introductions and conclusions. Postegullo linked such anomalies with the fact that computer science was then not yet a full-fledged discipline.

Kanoksilapatham contributed two studies to the literature on all-four section RA genre research (2007, 2015). In her 2007 study, she identified 12 moves in biochemistry RAs while she found 15 moves in her 2015 comparative study of three engineering sub-disciplines. Despite the difference in the number of moves, the rhetorical move structures were found quite similar. In the introduction section, both disciplines displayed the same three moves (named differently, though) predicted by the CARS model (Swales, 1990, 2004). In the methods section, four moves were identified in biochemistry but three in engineering. Of the three overlapping moves,

the *Reporting and consolidating findings* move is quite unusual given that her corpus contained exclusively artilces with stand-alone methods and results sections. Kanoksilapatham regarded this as evidence of the gradual fusion of the methods and results sections in the three engineering sub-disciplines. As to the results section, four moves were present in biochemstry RAs whereas engineering RAs had three, of which the *Summarizing procedures* move was actually a combination of two separate moves in the biochemstry RAs. Finally, the two disciplines were found to be similarly structured in the discussion sectiion except that the two moves of *Stating limitations of the study* and *Suggesting further research* in biochemistry were combined into a single move, *Stating limitations and future research*, in engineering. Two striking charateristics of Kanoksilapatham's studies are rigorous inter-coder reliability check and inferential statistics used to establish significant differences. Of particular interest is her 2007 study that innovatively used MD analysis to characterize rhetorical moves in terms of how they used different sets of co-occurring linguistic features to realize their communicative purposes. This part of the study will be reviewed in detail in Subsection 2.5.2.

Similarly, in a fairly large corpus of chemistry RAs, Stoller and Robinson (2013) found two predominant organizational patterns: IMR[DC] and IM[R(DC)], with brackets indicating sections merged under one major heading. With respect to rhetorical move structure, chemistry RAs seemed to have a less complicated structure

in that ten moves were identified. Interestingly, for pedagogical purposes, the researchers organized the moves into easy-to-interpret instructional tools labeled ''move structures akin to flow charts.''

One of the rare published studies of all-four section RAs in social sciences and humanities is Pho's (2008a) comparative study of applied linguistics and education technology RAs. Pho found that in general, applied linguistics RAs and education technology RAs were similarly structured. Introductions in both disciplines employed the same three moves in the CARS model (Swales, 1990, 2004). In the other sections, applied linguists made a total number of eleven moves while education researchers made nine, which were employed by all applied linguists. Pho also observed that applied linguistics RAs were more elaborate than education technology RAs. In particular, in the methods section, applied linguistics articles tended to follow up recounting the data collection procedure with a justification. Similarly, in the results section, applied linguists did not only report their findings but made an extra move to comment on them. As for the discussion section, applied linguistics RAs tended to make two extra moves that were found infrequent in education technology RAs: *preparing for the presentation of the discussion by giving background information*, and *drawing conclusions of the study*.

Another study is the most recent Tessuto (2015) on empirical law RAs. The study revealed marked deviations of law RAs from the conventional IMRD organizational

pattern. A Background Review and a Conclusion section were found prevalent in the articles. Background Review elaborated on the key points outlined in Introduction by situating the research problem in a wider and more detailed context, thus complementing Introduction which presented only a general picture of the research. The Conclusion section returned to the research question raised in Introduction by associating the narrow data analysis with the broader concerns with which the article began. Accordingly, in analyzing the generic move structure, Tessuto treated Background Review as independent of Introduction and Conclusion as independent of Discussion. The move structure of Background Review was found to be identical with that of Introduction. While the Conclusion section shared with Discussion two moves, viz., *Evaluating the study* and *Deductions from the study,* it did not provide background information or reinforce the results as the Discussion did.

In sum, all-four section RAs research has revealed much disciplinary variation in terms of both sectional structure and rhetorical move structure. The present study drew on this research as well as the research on individual sections in analyzing the rhetorical move structure of education RAs. In particular, Pho's (2008a)14-move model contributed most to the initial coding scheme for the move analysis because of its comprehensiveness in rhetorical structure characterization and its relevance to the general field of education.

2.3 Corpus Linguistics and Genre Analysis

Modern corpus linguistics came into being with the emergence of the first computer corpus of the English language, the Brown Corpus, in the early 1960's. Ever since, and particularly since the 1980's, it has flourished with an increasing number of corpora of different kinds and sizes, utilizing more and more sophisticated computer software programs for analysis. Applications of corpus technologies are found in virtually all fields related to language such as language teaching and learning, lexicography, grammar, discourse analysis, and translation studies.

2.3.1 Overview of Corpus Linguistics

Central to corpus linguistics are collections of authentic texts called corpora which serve as empirical data for studying language as it is used in real life. Sinclair (1991) defined a corpus as "a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language" (p. 171). His definition captures the two most important properties of corpora: authenticity of language and representativeness of text samples. However, he did not include computer-readability in this definition because it is a standard property of modern corpora. McEnery and Wilson (2001) described a modern corpus more accurately as "a finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration" (p. 32).

Definitions of corpora are many, but it is generally agreed within corpus linguistics that a corpus consists of computer-readable, authentic texts representative of

a particular language or language variety. Representativeness is an essential feature that distinguishes a corpus from a random collection of texts. Biber's (1993) discussion about representativeness is revealing. He defined representativeness in terms of how this quality is achieved, saying that representativeness is "the extent to which a sample includes the full range of variability in a population" (p. 243). Therefore, representativeness is, to a great extent, determined by size and the range of registers/genres included. The representativeness of a general corpus, which serves as a basis for an overall description of a language or language variety, depends on sampling from a broad range of registers/genres. In contrast, the representativeness of a specialized corpus is determined by the degree of "closure" or data saturation. A particular feature of a variety of language (e.g., RAs) reaches the point of closure when no further variation is indicated by the addition of more texts to the corpus.

Computer-readability involves two assumptions. First, texts are electronically stored so that they can be read by the computer. Second, computers and necessary software are used for processing texts and generating evidence of language in use, usually in the form of keywords lists and concordances.

Texts can be raw, but are increasingly annotated with tags assigned to lexical, syntactic, semantic and even pragmatic categories in them. As regards the value of corpus annotation, there are two opposing views. To those advocating a corpus-driven approach such as Sinclair (2004) and Tognini-Bonelli (2001), texts should be

minimally annotated to allow them to "speak for themselves." They maintain that

previous theoretical statements about language, for example, grammatical categories,

are problematic and therefore should not be used as a priori categories in corpus studies.

In contrast, corpus-based researchers view pre-existing linguistic categories as an asset

although they still need to be tested and revised. To them, annotated corpora have

"added value" because they not only facilitate automatic analysis, but, more

importantly, offer new ways of gaining greater understanding of language. Leech (1997)

claimed that annotation is "a crucial contribution to the benefit a corpus brings, since it

enriches the corpus as a source of linguistic information for further research and

development" (p. 2).

Corpus linguistics as a powerful methodology is well established. Compared

with the traditional intuition-based approach to language, corpus linguistics has the

advantage of improved reliability based on large amounts of empirical data. According

to Leech (1992) and Biber et al. (1998), corpus linguistics has, among others, these

important characteristics: 1) it represents empirical rather than rationalistic research, 2)

it is both quantitative and qualitative, 3) it studies natural texts, 4) it studies linguistic

performance rather than linguistic competence, and 5) it makes extensive use of

computer for data analysis.

Despite its power of processing large quantities of data of authentic language and

generating more reliable quantitative evidence based on frequency counts, corpus

linguistics is not free of weaknesses. Flowerdew (2005) summarized, among others,

two major disadvantages identified by several researchers. These disadvantages are

related to corpus-data presented in the form of concordances and keywords. First,

corpus-data are decontextualized. As pointed out by Widdowson (1998, 2000), corpus

linguistics describes text, not discourse, because corpus-data are removed from the

communicative context where they were produced. Second, corpus studies are limited

to a somehow atomized, bottom-up type of analysis which runs counter to the top-down

approach to genre analysis with a focus on larger units of text rather than sentence level,

lexico-grammatical features. A further disadvantage of corpus linguistics concerns

corpus size. To be representative, a corpus has to be sufficiently large (Sinclair, 1991,

2004). There is always a potential risk involved in relying on corpora that are not large

enough.

2.3.2 Interface between Corpus Linguistics and Genre Analysis

Due to their innate limitations, it seems that neither corpus linguistics nor genre

analysis alone can provide deep yet comprehensive descriptions of language use. As far

back as 1995, Gledhill (1995) observed that "hitherto genre analysis has not fully

capitalized on phraseology, and corpus linguistics has not begun to treat its texts

communicatively" (p. 12). The potential of a combined approach of both genre and

corpus analysis is obvious. Gledhill (2000) envisioned the complementary roles the two

approaches can play: while corpora will identify recurrent patterns of linguistic features

across a large number of texts, genre analysis will reveal contextual, ethnographic information on linguistic choices allowing us to sort out the subtle, often monolithic descriptions of corpus studies. In a similar vein, Biber et al. (2007) called for a merge of the analytical goals and methods of corpus linguistics with those of discourse analysis, suggesting that a corpus can be analyzed to identify the general patterns of discourse organization that are used to construct texts, and that individual texts can be analyzed in terms of the general patterns that result from corpus analysis.

To corpus-based genre studies, then, genre-specific corpora are the key. As general corpora are rarely subdivided strictly in terms of genres, they can not be adequately representative of specific genres (L. Flowerdew, 2004; Kaltenböck & Mehlmauer-Larcher, 2005). Biber (1990), Flowerdew (2004) and Tribble (1997, 2001) advocated the use of small, subject-specific and goal-oriented corpora. Gavioli (2005) stated that specialized corpora, by way of concordances, reveal the typical, conventional ways that lexis and textual or generic structures are used inside specialized texts.

The next section is devoted to previous corpus-based genre studies of RAs.

2.4 Corpus-based Research on RA Linguistic Features

In recent years, the intensive manual labor on a small set of texts that used to characterize genre research has started to give way to automatic analysis of large samples of authentic texts. Empowered with corpus techniques, ESP/EAP researchers

have become able to describe lexico-grammatical patterns in RAs that were previously

unobservable. Their research can be grouped into two major categories: studies of

phraseologies, and studies of metadiscourse, although they may overlap.

Before reviewing corpus-based research, it is necessary to do justice to earlier

non-corpus based studies by summarizing the important ones.

2.4.1 Non-corpus-based Studies of Linguistic Features

Most of the non-corpus-based studies were concentrated in the 1980's and 1990's

before corpus linguistics had found much application in ESP/EAP genre research.

During this period, what attracted most of the attention was verb tense (e.g.,

Gunawardena, 1989; Gerbert, as cited in Holtz, 2011; Trimble & Trimble, 1982) and

voice (e.g., Tarone, Dwyer, Gillette, & Icke, 1981; Wingard, 1981).

Studies on verb tense have shown that the present tense is usually used in

presenting definitions, descriptions, and observations, while the perfect tense is mostly

used in describing research processes (Gerbert, as cited in Holtz, 2011). It has also been

observed that the present simple and present perfect are often used to review previous

research in the introduction section and to discuss findings in the discussions whereas

the past tense is usually used in the methods and results sections to describe procedures

and report results (Gunawardena, 1989; Trimble & Trimble, 1982).

As to grammatical voice, Tarone, Dwyer, Gillette, and Icke (1981) found that the

active voice is often used to describe the author's own work whereas the passive voice

is more associated with previous studies. Another interesting study revealed co-occurrence patterns of voice and tense (Wingard, 1981). The passive voice was found to work in tandem with the past tense in the methods section to describe the procedures of the current research, while the active voice in the present tense describes an apparatus. In another study, Riley (1991) linked passive structures with expository purposes and active structures with argumentative purposes in different RA sections.

All these studies analyzed authentic, but typically few texts. Biber, et al. (1998) pointed out that "they were not typically corpus-based inveatigations…[and] few of these studies aim to produce generalizable findings that hold across texts" (p.106). However, a pronounced shift towards using corpus methods in genre research has been evident in genre research since the turn of the 21st century.

2.4.2 Studies of Phraseologies in RAs

Corpus-based lexico-grammatical studies look for reccurrent patterns in language use. This explains why research on lexico-grammatical features of RAs are characterized by extensive interest in phraseologies including collocations and multiword expressions. Collocations are often understood as pairs of words that tend to co-occur more frequently than expected by chance (Hunston, 2002). Multiword expressions are extended collocations. They are also known by various names such as "clusters" (Hyland, 2008a), "lexical bundles" (Biber & Barbieri, 2007), and "n-grams" (Stubbs, 2007), but basically these terms refer to those recurrent word sequences of three or more words retrieved from a corpus with specified frequency and distribution

criteria. Collocations and multiword expressions are thought to have customary pragmatic and discourse functions, used and recognized by the discourse community.

Two examples of studies on collocations in RAs are Williams (1998) and Luzon Marco (2000). Williams (1998) compared collocations used in RAs in molecular biology of parasitic plants and physiology of parasitic plants and those in papers presented by members of each discourse community who met at a conference to form a third community with an interest in a common topic. He first built a network of words associated with one another by identifying the collocates of the high frequency words in each corpus and then the collocates of those collocates. These networks were then compared between the discipline-based discourse communities and the topic-based discourse communities. The collocations in the topic-based corpus reflected the varied concerns and methods of the different discipline discourse communities. Unlike Williams who investigated continuous collocations, Luzon Marco (2000) examined the most frequently occurring collocation frames in a corpus of medical RAs. He found, for example, that the frame *the ...of* had some specific functions in medical RAs. One of these was to express measurement and quantification, as in *the amount of*, *the degree of*, *the extent of*. Another function was to express a focus on an aspect of the research, as in *the cause of*, *the effect of*, and *the risk of*.

Most of the studies on multiword expressons in RAs have taken a comparative perspective. For example, Cortes (2004) compared lexical bundles as used by professional writers and student writers in two disciplines: history and biology. She

found that students rarely used the bundles identified in professional writing. Furthermore, when students did employ certain bundles in their writing, they used them differently than professional writers. In another study of RAs in historical research across English and Spanish, Cortes (2008) found that the bundles identified in each language had many features in common. While one group of bundles seemed to be the result of direct translation, a second group of bundles showed structural characteristics that were closely related to bundles frequently found in academic writing in both languages. Finally, in terms of function, some bundles from the two languages demonstrated similarities shared by all academic prose, by historical research, and by similar topics across the journals.

Hyland (2008b) explored disciplinary variation by investigating the forms, structures and functions of 4-word bundles in a 3.5 million-word corpus of academic prose, including research articles, doctoral dissertations and Master's theses in electrical engineering, microbiology, business studies and applied linguistics. His results showed that bundles are not only central to the creation of academic discourse, but that they offer an important means of differentiating written texts by discipline. Hyland concluded that disciplines differ in the frequencies of use of bundles and in the principal structures of bundles used.

In another study, Gledhill (1995) investigated lexical features in relation to different sections of RAs, including abstracts. His corpus was a highly specialized corpus of RAs on cancer research. Based on the keyword list technique, he found, for

example, that the methods and results sections contained verbs of research process such

as *added*, *performed*, *incubated*, *obtained*, and *dried*; verbs of observation such as

shown, *described*, *reported*, *expected*; and specific methodological terms such as

solution, *temperature*, and *buffer*. By comparison, in the introduction and discussion

sections, the salient items were auxiliaries, including modals, such as *is*, *would*, *might*;

and items which signal cohesion and the logical connection between one part of the

discourse and another, such as *because*, *then*, and *each*. On the other hand, his

concordances revealed salient grammatical items that demonstrated lexical patterning

in the different sections of the RAs. For example, in the discussion section, one frame

involving the preposition *in* is *play a ... role in*. The noun *role* is modified by an

adjective such as *major*, *important*, and there is often modality in the sentence. This

suggests that the frame expresses the author's judgment.

Apart from a focus on phraseologies, there have been studies on such features as

reporting verbs and citations (Martínez, 2008; Thomas & Hawes, 1994; G. Thompson

& Ye, 1991; P. Thompson & Tribble, 2001), lexical verbs (I. A. Williams, 1996), and

discipline-specific vocabulary (Martínez, Beck, & Panza, 2009).

2.4.3 Studies of Metadiscourse in Academic Writing

Metadiscourse is based on the view of writing as a social engagement. Rather than

merely providing information objectively, academic writers personally interact with the

reader by guiding them through the text and influencing them with his or her own attitude

towards the propositional content and the reader. Metadiscourse in RAs as well as in other types of academic writing has generated intense research interest in recent years (e.g., Baratta, 2009; Gillaerts & van de Velde, 2010; Hewings & Hewings, 2002; Hu & Cao, 2011; Hyland, 2005b; Hyland & Tse, 2004, 2005; Lewin, 2005; Mur-Dueñas, 2011). The purpose of metadiscourse research is to reveal how writers intervene in the text to evaluate their knowledge claims, comment on them, and build solidarity with their readers. Some metatdiscourse studies have investigated a category of metadiscourse or metadiscourse in general; more have focused on a specific metadiscourse device.

Hyland is the most productive researcher on metadiscourse (1996, 1998a, 1999, 2000, 2002a, 2005a, 2005b, 2008c; Hyland & Tse, 2004, 2005). His work has expanded our understanding of metadiscourse in RAs. He has extended the term *metadiscourse* from its original concept of "discourse about discourse" to cover "the linguistic resources used to organize a discourse or the writer's stance towards either its content or the reader" (Hyland & Tse, 2004, p. 157). In his model of metadiscourse in academic writng, metadiscourse is represented by interactive and interactional linguistic resources or devices (Hyland, 1998b, 2005a). Interactive devices function to guide the reader through the text. The author proposes an argument to explicitly establish his or her own preferred interpretations thus influencing the reader's understanding. Differently, interactional metadiscourse engages the reader in the argument by signaling the author's attitude towards both propositional information and the reader.

One example of comprehensive analyses of metadiscourse is Hyland's (1998a) analysis of a corpus of 28 RAs in four disciplines. This study demonstrated that, in general, metadiscourse has high frequencies of occurrence in RAs. The quantitative results suggested that academic writers use more interactive (textual) devices in RAs than interactional (interpersonal) devices and that hedges (e.g., *might/perhaps/about/it is possible*) and connectives (e.g., *in addition*, *but*, *therefore*) are the most frequently used, followed by code glosses (e.g., *namely*, *such as*, *in other words*) and evidentials (e.g., *according to X & Y (1990)*, *Z states*). The predominance of interactive metadiscourse emphasized the common understanding of metadiscourse as guiding the reading process by signposting discourse organization, and revealing propositional connections and meaning. Specifically, the most useful interactive device is logical connectives. The study also revealed that hedges are by far the most useful interactional device. This finding highlights the importance of distinguishing facts from opinions and the need for RA writers to present their knowledge claims with sufficient caution. In terms of the comparison among the different disciplines, despite the similar density of metadiscourse, substantial differences were found. For example, while hedging is the most freqently used device in nearly all disciplines, it is second to connectives in astrophysics.

In contrast to Hyland's 1998 study of all-four-section RAs, Abdollahzadeh (2011) restricted his study to the Conclusion section. He investigated hedges, emphatics

(boosters) (e.g., *in fact*, *definitely*, *it is clear*), and attitude markers (e.g., *surprisingly*, *I agree*, *X claims*) in a corpus of 60 applied linguistics RAs written in English by Anglo-American and Iranian academic writers. Both groups exhibited a strong tendency towards hedging their propositions. However, Anglo-American authors were found to use more emphatics and attitude markers than their Iranian counterparts. The researcher attributed Iranian writers's limited use of attitudinal language to their higher professional and academic status than that of the local audience their articles were targeted at.

Other studies of a range of metadiscourse devices include, to name a few, Hyland (2005b) under the heading of stance and engagement; Pho (2010) on authorial voice realized by self-reference words, stance expressions, modal auxiliaries and semi-modal verbs; Gillaerts & van de Velde (2010) investigating interpersonality in terms of hedges, boosters and attitude markers; and Dahl (2004) comparing interactive (textual) metadiscourse in three languages across three disciplines.

While research on an entire category of metadiscourse provides a general view of social interaction involved in RAs, studies on specific metadiscourse devices offer deep insights into the specific ways of realizing this social interaction. Research on individual metadiscourse devices include studies on hedging (e.g., Hyland, 1996; Lewin, 2005; Salager-Meyer, 1994), emphatics (boosters) (e.g., Bondi, 2008), self-mentions (e.g., *I*, *the researcher*, *the author*) (e.g., Mur-Dueñas, 2007a), logical

markers (e.g., Mur-Dueñas, 2007b), attitude markers (e.g., Mur-Dueñas, 2010). Besides, included in this category are also studies on specific lexical and grammatical items that express the author's stance, for example, anticipatory *it* (Rodman, 1991), personal pronouns (Kuo, 1999), and the passive voice (Baratta, 2009).

2.5 Multidimensional Analysis

Within corpus linguistics, multidimensional (MD) analysis was first used by Biber (1986, 1988) to achieve a comprehensive characterization of the patterns of systematic register variation in English. Ever since, MD analysis has been extended to a wide range of research areas of linguistic variation including, for example, synchronic analyses of specific registers and genres and author styles; diachronic studies describing the evolution of registers; register studies of Non-Western languages and contrastive analyses; research of university English and materials development; and move analysis and study of discourse structure (see Xiao, 2009; Xiao & McEnery, 2005).

In the following subsections, Biber's MD analysis framework (1986, 1988) is described, followed by a focused review of Kanoksilapatham's (2007) MD analysis of biochemistry RAs.

2.5.1 Biber's MD Analysis Framework

MD analysis is a corpus-based quantitative approach to the analysis of linguistic variation among registers or genres. The method uses the statistical procedure of factor

analysis to identify patterns of co-occurrence of a broad range of linguistic features in text thus revealing the major dimensions on which texts vary.

MD analysis assumes that some linguistic features tend to occur together because they share similar or closely-related functions which can be grouped together for clearer understanding. For example, first- and second-person pronouns, direct questions, and imperatives tend to co-occur in texts because they all relate to interactiveness. A set of co-occurring features is called a dimension of variation. Another important assumption is that because differences among registers/genres are hardly absolute or clear-cut, other than multiple dimensions, no single dimension can accurately measure linguistic variation among them (Biber, 1988). For example, in his study of a corpus of 481 texts across 23 spoken and written English genres, Biber (1988) identified six dimensions: 1) Informational vs. involved production, 2) Narrative vs. non-narrative concerns, 3) Explicit vs. situation-dependent reference, 4) Overt expression of persuasion, 5) Abstract vs. non-abstract information, and 6) On-line information elaboration. As can be seen, a dimension typically consists of two sets of features, which are in a complementary relationship. That is, when the frequency of co-occurrence of one set is significantly high, the co-occurrence frequency of the other will be markedly low.

Two major components are included in an MD analysis: 1) identification of the underlying linguistic parameters, or dimensions, of variation; and 2) specification of the linguistic similarities and differences among registers with respect to those

dimensions. Methodologically, MD analysis has three defining characteristics: 1) an automatically tagged representative corpus; 2) a concordance program for obtaining frequencies of linguistic features in texts; and 3) computerized factor analysis to analyze the co-occurrence relations among a large set of linguistic features, thereby uncovering dimensions of variation.

MD analysis has proved to be a valuable analytical framework for both inter-textual and intra-textual linguistic variation (Biber, Conrad, Reppen, Byrd, & Helt, 2003; Getkham, 2010; Kanoksilapatham, 2007; Xiao, 2009; Xiao & McEnery, 2005). Its success can be attributed to three major strengths associated with it. First, by taking the advantage of corpus-based research methodologies to investigate overall patterns of linguistic variation on the basis of a representative sample of texts, this approach yields more generalizable results. Second, computer programs and interactive software tools make it possible to analyze a large number of linguistic features in the texts. And third, relying on factor analysis, it can deal with a large number of independent variables, for example, the different moves in the case of this present study.

Nevertheless, MD analysis is limited to linguistic features within the sentence, thus unable to provide a description of the macro-organizational structure of a genre. It is therefore best to integrate it with genre analysis that focuses on textual chunks that may be larger than sentences.

2.5.2 Kanoksilapatham's MD Analysis of RAs

Using corpus-based multidimensional (MD) analysis developed by Biber (1985, 1986, 2008), Kanoksilapatham's dissertation (2003, published later in article form in 2005 and 2007) investigated how 41 lexico-grammatical constructions co-occurred in different combinations to convey the communicative purposes of the rhetorical moves identified across all four sections of sixty biochemistry RAs

Kanoksilapatham (2007), based on her 2003 doctoral dissertation, was the only published investigation of RAs to follow up a move analysis with an MD analysis of linguistic variation among the rhetorical moves. The MD analysis investigated how different sets of linguistic features realized the communicative purposes of the moves identified in a corpus of 60 biochemistry RAs. The results showed seven underlying dimensions of linguistic variation: 1) Conceptual vs. concrete feference, 2) Concrete action vs. abstract discussion, 3) Evaluative stance, 4) Projected interpretation, 5) Attributed knowledge vs. current Study, 6) Stated purpose, and 7) Contradictory proposition. These dimensions accounted for 33.5% of the total linguistic variance among the moves, an acceptable result considering the homogeneity of the genre investigated. Each move was measured by all seven dimensions and the readings of the move on all the dimensions culminated in a full profile of the move in terms of the extent to which it is marked by the different sets of co-occurring linguistic features. Subsequent ANOVA tests based on the mean dimension scores of the moves indicated

significant differences among the 13 moves (2 were excluded due to their limited number of texts) with respect to each of the seven dimensions.

Kanoksilapatham's (2007) study was a groundbreaking experiment in the integration of corpus linguistics and genre analysis. The innovative use of Biber's (1986, 1988) MD analysis to analyze internal discourse structure demonstrated that corpus linguistics can be successfully extended to genre analysis, and that MD analysis is not only effective in analyzing register variation, but it can be equally useful in revealing linguistic variation between different communicative chunks of text. The combined use of the two approaches can generate more comprehensive descriptions of the ways texts are constructed in terms of communicative purpose and linguistic choice.

However, Kanoksilapatham's (2007) study could have been more successful but for the problem of data loss. For statistical reasons, all texts of two moves with limited observations and all move texts of 25 or fewer words had to be excluded from the MD analysis, resulting in a loss of as much as 28% percent of the corpus data. This problem could have been mitigated if move-embedding had not been considered in the move analysis, as embedded moves are normally short phrases within sentences.

The present study represents another attempt to investigate inter-move variation of the research article genre using MD analysis.

2.6 Critique of Previous Research on RAs

Previous research has provided invaluable insights into the RA genre, both beyond and within the sentence level. Swales' move analysis framework (1981, 1990, 2004) has empowered ESP/EAP researchers to describe how the overall communicative purpose of the RA genre can be realized by different moves and their constituent steps. The numerous studies of the macro rhetorical structure of RAs have captured the differences between disciplines, sub-disciplines, and languages. While move analyses have focused on the rhetorical structure, corpus-based studies focusing on sentence-level linguistic choices in RAs have also enhanced our knowledge of expert writers' linguistic choices in relation to communicative purposes.

However, despite the achievements made so far, a few limitations of past research have emerged from this review. Some are related to oversight on the part of researchers, and some are related to the methodological complexity of RAs.

Human oversight is reflected in a narrow focus of both genre research on the RA rhetorical structure and corpus research on RA linguistic features. Within genre research, while extensive attention has been accorded to natural sciences, limited research has been conducted on social sciences and humanities genres. Only a few disciplines of social sciences and humanities, for instance, applied linguistics, have been investigated so far. Likewise, in contrast to the overwhelming number of comparative studies aimed at uncovering the differences in rhetorical structure across

and within disciplines, studies are rare on the rhetorical organization universal in a range of inter-related disciplines, for example, education research that is at the interface of education and many other academic disciplines. General descriptions of the rhetorical structure, like Swales' CARS model (1981, 1990, 2004), would be just as helpful to ESP/EAP instruction. The assumption is that inter-related disciplines are expected to be much more alike than different. Pedagogical needs should also be considered—few ESP/EAP classes are made up students from a single discipline or field. An additional limitation associated with genre analysis is the clearly much less attention paid to the other sections of the article than the introduction and discussion sections.

Corpus studies of linguistic features in RAs have the virtue of being able to provide some concrete descriptions of the RA genre by showing how lexico-grammatical items behave in the articles. Yet the features investigated in separate studies are often highly discrete from one another—that is, they constitute an array of fragmented indicators of the communicative purposes that writers wish to fulfill. As such, they are difficult to synthesize into a very small number of indexes of communicative functions.

Attempts to provide a more complete picture of the RA genre through examining an aggregated number o features have been made. Nevertheless, studies such as Hyland's (2005b) examination of a broad range of linguistic features are rare,

and Kanoksilapatham's (2007) MD analysis of a large set of linguistic features in relation to rhetorical moves has remained exceptional. Biber's (1986) call for macro investigations into the association patterns of different linguistic features are yet to be answered regarding the rhetorical moves in RAs.

Methodologically, the qualitative nature of move analysis has given rise to two marked problems with this analytical approach. The first has to do with the subjective nature of move identification (Kanoksilapatham, 2007). Move analysis tends to be subjective because the analyst usually relies on his or her intuition for move coding (Crookes, 1986; Paltridge, 1994), and different analysts may assign different moves to the same text.

The second major problem frequently associated with many genre studies is related to the sampling bias involved. This view corroborates Biber et al.'s (1998) observation that unlike corpus studies, many discourse studies have been based on small sets of texts. Indeed, most of the corpora used in previous genre studies ranged from 10 to 60 texts, and decisions on corpus size has been rather intuitive and arbitrary (e.g., Brett, 1994; Henry & Roseberry, 1999; Samraj, 2002). Besides insufficient sample size, Kanoksilapatham (2007) points out that unrepresentative data often result from subjective selection of journals based on mere experts' recommendations (e.g., Nwogu, 1997; Posteguillo, 1999) and a lack of specification criteria for text selection (e.g., Swales & Najjar, 1987; I. A. Williams, 1999). Problematic samples jeopardize

any claim of reliability and findings about the rhetorical organization and lexico-grammatical features can by no means be taken as conclusive.

2.7 The Present Study

The gaps in previous research literature on the RA genre and the need of NNS education researchers for assistance in research article writing suggest that it would be a worthwhile attempt to conduct a comprehensive investigation into education RAs. Therefore, the present study was aimed at investigating both the overall move structure of education RAs and language use in relation to the rhetorical moves. The research problem was approached from a double perspective of genre analysis and corpus linguistics. Two objectives were attempted: 1) to identify the rhetorical move structure of education RAs; and 2) to map out linguistic co-occurrence patterns typically associated with the communicative purpose of each move in education RAs.

2.8 Summary

This chapter examined the theories that inform the present study, including genre theory, particularly Swales' move analysis, and corpus linguistics, particularly Biber's MD analysis. Meanwhile, previous research on the RA genre from the two perspectives of genre and corpora was reviewed, though by no means exhaustively. While the review of relevant theories highlighted the advantage of a corpus approach to genre analysis,

the examination of past research helped pinpoint some deficiencies of research on the RA genre that need to be addressed. The review arrived at a conclusion that a mixed methods inquiry into the education RA genre as a whole would make a useful contribution to ESP/EAP research and practice.

Chapter 3 will elaborate on the methodology of the present study.
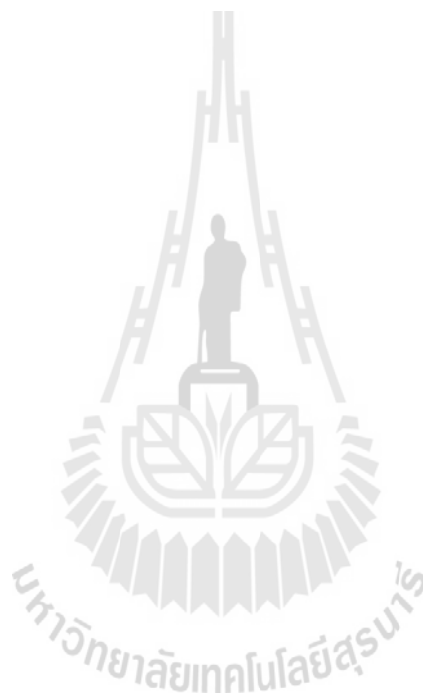
CHAPTER 3

METHODOLOGY

This chapter describes different aspects of the methodology of the study. Section 3.1 introduces the sequential mixed methods research design that the present research employed. Section 3.2 elaborates on the construction of a specialized corpus that served as data input for the investigation. Descriptive details are provided with regards to text sampling, text annotation, and the derivation of sub-corpora of moves. The other major aspect of the methodology, data analysis, is dealt with in two separate sections. Section 3.3, which is devoted to move analysis, covers the analytical framework, move identification criteria, inter-coder reliability, and the derivation of a prototypical move structure. In subsequent Section 3.4, the MD analysis framework is introduced, followed by a description of the procedures of statistical analysis.

3.1 Research Design

To address the research questions, a two-phase sequential mixed methods design was developed for the present empirical study of education RAs. Figure 3.1 illustrates the components of the research design and the major procedures involved in executing the research. In the first phase, genre analysis within the framework of Swales' move analysis (1990, 2004) was employed to analyze the rhetorical move structure of

education RAs as represented by the Education Research Corpus (ERC) especially

constructed for this study. In the second phase, a corpus approach was taken to

investigate the linguistic features of the articles. For this purpose, Biber's (1986, 1988)

MD analysis was employed to analyze the typical ways linguistic features combine to

realize those rhetorical moves.
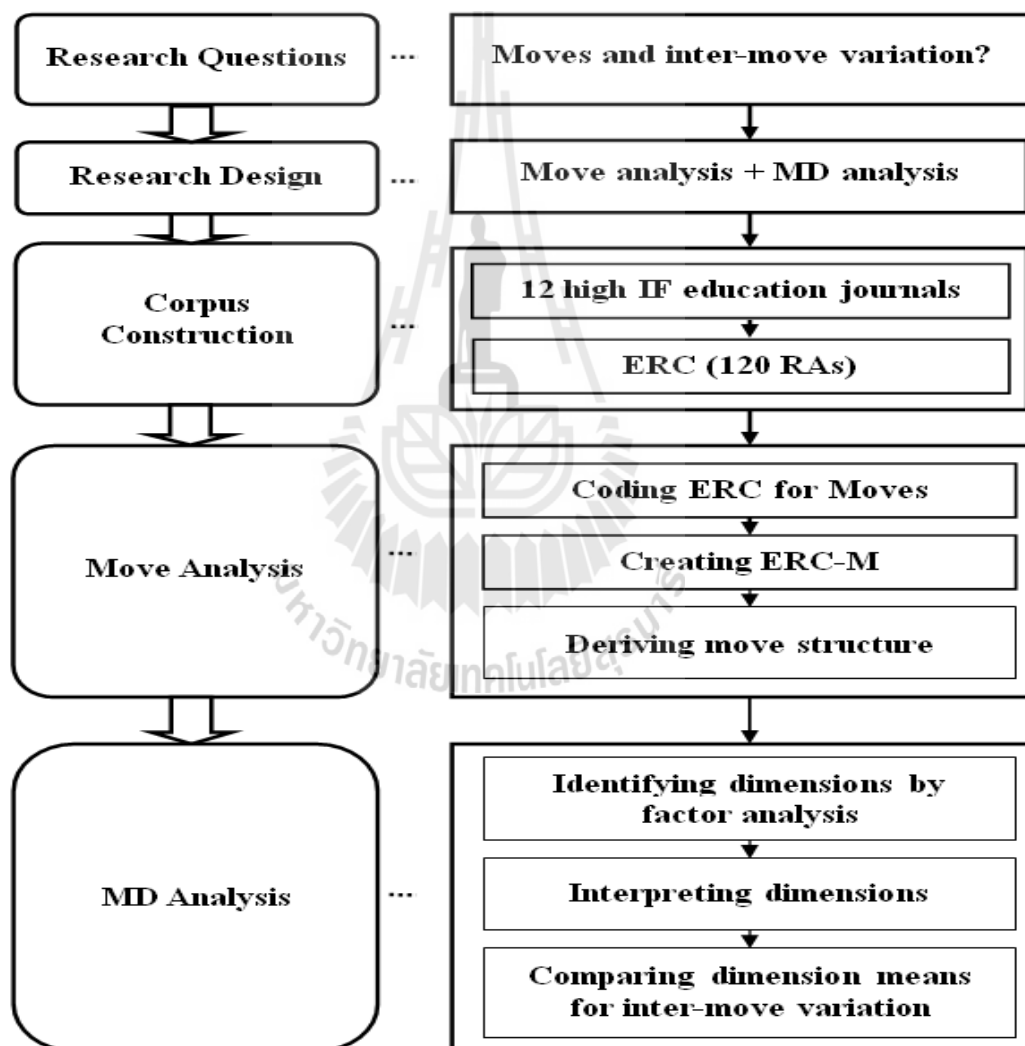
Figure 3.1 Research Design Flow Chart

Legend: MD analysis = multidimensional analysis; ERC = Education Research Corpus;

ERC-M = Education Research Corpus of Moves

Methodologically, this investigation is characterized by a triangulation of the primarily qualitative move analysis and the primarily quantitative corpus approach to analyzing the move structure and linguistic features. The advantage of the combined use of multiple methods in examining different aspects of the same phenomenon (in this case, education research articles) lies in the enhanced validity of research results that constitute a holistic and rich account of the phenomenon (Scott & Morrison, 2005). As discussed in Chapter 2, the complementary roles that genre analysis and corpus analysis can play has been acknowledged by several researchers. Relying on the genre-specific, goal-oriented ERC corpus, a move analysis identified the rhetorical move structure of education RAs while an MD analysis revealed the co-occurring lexico-grammatical patterns across the rhetorical moves. That is, a move analysis addressed Research Questions 1 and 2 while the results of an MD analysis answered Research Questions 3 and 4.

3.2 Corpus Construction

3.2.1 Corpus Size Considerations

A specialized corpus of education RAs named the Education Research Corpus (ERC) was constructed for the purpose of investigating the target genre.

Although it was impossible to determine a prior the minimum size of a corpus for it to begin to be representative, an estimate was made on the basis of the number of

rhetorical moves anticipated and the number of linguistic features under examination. As Flowerdew (2004) points out, the appropriate size for a specialized corpus is highly dependent on the linguistic phenomenon to be investigated. For the sole purpose of analyzing the rhetorical move structure of a genre, a corpus of a few dozen texts may be sufficient because of the high frequencies of occurrence of the moves in the texts. By comparison, arguably a much larger corpus is needed if a broad range of linguistic features are to be analyzed. This is so because while common features have high frequencies of occurrence, some relatively rare ones require a larger amount of text to provide sufficient information on them.

Previous multidimensional studies of linguistic variation adopted a minimal 5:1 ratio of number of texts (in this case, texts of rhetorical moves) to number of linguistic features to be analyzed (Biber, 1988; Kanoksilapatham, 2007). To arrive at an approximate estimate of the minimum number of articles needed to generate valid results, a preliminary survey was conducted of articles published in the data source journals. It showed that education RAs tend to have greater than average lengths. The average number of instances of a move in education research article seemed likely to exceed three. Suppose that three was the maximum average. Given that the number of linguistic features to be analyzed in the study was 67, then the number of articles required was calculated thus:

Number of articles = $(67 \times 5) \div 3 = 111.7$

Therefore, in order for the corpus to be adequately representative of education research thus ensuring validity of results, a total of 120 empirical education RAs was systematically sampled from 12 journals, i.e., 10 articles from each. The corpus was expected to be approximately a million running words in size, sufficiently large to allow for an analysis of both rhetorical moves and a broad range of linguistic features realizing the moves.

3.2.2 Data Collection

This subsection describes the criteria and procedures for selecting journals and articles for the ERC which served as primary data for the study.

3.2.2.1 Selection of Journals

In selecting journals as data sources, the three general criteria set by Nwogu (1997), namely, representativeness, reputation and accessibility, were observed. The majority of the previous studies had suffered from a sampling bias resulting from using whatever journals recommended by experts (e.g., Hyland, 1999; Nwogu, 1997; Posteguillo, 1999). Kanoksilapatham (2007) argued that expert recommendations could be subjective in that they might reflect individual preferences. To enhance objectivity, researchers (e.g., Kanoksilapatham, 2007; Parkinson, 2011; Pho, 2008a) are increasingly choosing high impact factor (IF) journals from which to draw research articles. Selecting journals according to IFs would result in a corpus considerably more representative than many corpora previously used in similar studies.

Therefore, impact factors (IF) were used in this study as a basis for journal selection. The "Top Journals in Education Ranked by 2007 Impact Factor" list published by Thomson Reuters (2009) included 20 journal titles that are considered the world's most influential and prestigious in the field of education research. Available databases provided full-text download service for 15 of the 20 titles. In order for the corpus to be more balanced, three journals, one on AIDS education and two on literacy education, were dropped from the 15 titles. The journal on AIDS education was excluded on the grounds that there were two on health education and that it was too specialized. The exclusion of the two literacy education titles was due to the fact that there were three titles on the same subject.

Together, the 12 journal titles (see APPENDIX A) from which articles were selected for the ERC had a broad coverage of areas or fields of education research. As can been seen, 10 of the selected journals each had a focused area of education, including engineering education; management and business education; literacy education; health education; education technology; learning, development and teaching; educational sociology; applied linguistics; science education; learning theory and pedagogy. The other two titles (*American Educational Research Journal* and *Elementary School Journal*), not restricted to a specific academic field, further contributed to the breadth of disciplinary coverage of the corpus. Presumably, these 12 journals represented quality education research writing. In addition, they were believed

to reflect considerably more inter-journal variation than those previous corpus-based studies that sampled only a very limited number of journals.

3.2.2.2 Selection of Texts

Ten full-length articles were drawn from each of the 12 journals, resulting in a set of 120 texts. A survey of the 12 titles showed that, for the year 2011, full-text download service was provided for only 7 titles, and that for 2010, all journals but one could be accessed for full-texts. The one exception was *American Educational Research Journal* whose most recent downloadable issue was published at the end of 2009. Therefore, to control for possible chronological variation, articles were sampled from the 2010 issues of the 11 journals and the 2009 issues for the remaining *American Educational Research Journal*.

Ten articles were systematically sampled from each journal. Starting from the last article in the last issue and moving backwards, every *K*th article was selected until 10 articles were obtained from each. The interval size *K* varied from journal to journal and was computed by dividing the total number of articles by 10. Articles were selected from those that met the following criteria:

1) Only complete articles with a minimum length of 5,000 words were selected. Articles of a limited length might contain many rhetorical moves that were too short so that a sub-corpus of a particular move would be too small, thus skewing statistical analysis.

2) Following all previous research, article selection for the this study was restricted to empirical studies, i.e., those that used data derived from actual observations or experiments. Articles were drawn only from under the headings of "Articles," "Research articles," "Original articles," "Original contributions" or "Papers" in the selected sources. Theoretical articles and articles published in special issues were excluded. This was because the rhetorical structure of an article may vary in accordance with its type (Crookes, 1986; Ozturk, 2007).

3) Only articles conforming to the IMRD organizational structure or an identifiable close variant of the structure were legitimate for inclusion in the corpus. IMRD is the widely accepted conventional structure of empirical RAs. A preliminary survey of the selected journals showed that apart from those with a distinct IMRD structure, some articles had a structure that was a close variant of the standard structure but labeled particular sections differently. Articles thus structured were included as long as different sections could be identified by content. Specifically, besides those with a distinct IMRD structure, these types of articles were also included if they otherwise met the criteria: a) articles with an unlabeled introduction; b) articles with a methods section labeled differently as "Methodology," "The Present Study," "Procedures," etc.; c) articles having a combined Discussion and Conclusion section; and d) articles having a stand-alone Conclusion section. In the case of articles with an independent Conclusion section after the discussion section, the Conclusion sections

were combined with the discussion sections, following Holmes' (1997) and Pho's (2008a) practice. However, articles with a combined Results and discussion section were not considered, as the initial survey indicated that less than 5% of the articles were so structured.

It should be noted that no effort was made to check whether the authors of the articles were native or non-native speakers of English. This was because this study did not take native language status as a variable. Another reason was that there was no reliable way to do so. An author's name and institutional affiliation would not be reliable clues. Most importantly, all the articles published in the 12 highest impact factor journals in the field of education had all undergone a rigorous peer review, proof-reading, and editorial process and were therefore assumed a representative sample of expert writing (Lee & Chen, 2009).

3.2.3 Corpus Annotation

While it is possible to extract certain linguistic information present in a raw corpus, it is often necessary to annotate the corpus so that a broader range of linguistic features can be extracted and analyzed. The process of "adding such interpretative, linguistic information to an electronic corpus" is called annotation (Leech, 1997, p. 2). Annotation adds a new layer of information which can be counted, sorted, and compared. However, in this study the term "annotation" is used in an extended sense to refer to both the addition of linguistic and contextual information to the texts in the

corpus. Three types of annotation were performed: metadata markup, move tagging, and Part of Speech tagging.

3.2.3.1 Text Editing

The first step in annotating the corpus was to clean up the text files converted from the PDF files downloaded from the databases. This involved a) removing all elements irrelevant to the study including abstracts, key words, references, footnotes, acknowledgments, graphics, tables, figures and page numbers; b) deleting empty lines, redundant spaces, hard carriage returns, etc., as they were likely to cause tagging errors which might in turn lead to inaccurate analysis; c) standardizing spellings to ensure that the same words were spelt the same consistently to avoid errors in frequency counts; d) replacing mathematical formulas and foreign characters with the codes <MF> and <FW> respectively as they might cause tagging errors; and e) deleting extended quotes by participants in the results section. These quotes, normally colloquial in style and often extended in length, did not reflect the authors' language use, which was the very target of investigation of the present study. However, to retain the communicative roles they played in the articles and preserve the integrity of the sentence structure, the code <EX> was used to occupy their original places.

3.2.3.2 Move Markup

Move markup was carried out after the paper-based move analysis (see 3.3) was completed with text segments marked up in hand-written codes for moves and

steps. Adding move codes as tags to text segments in the .txt files served two purposes:

to enable automatic frequency counts of moves and steps in each section, and to enable

automatic creation of sub-corpora of the move texts drawn from the corpus as

necessitated by the task of investigating the linguistic features of the rhetorical moves.

Move markup was done semi-automatically using the BFSU Qualitative

Coder 1.0 (Xu & Jia, 2011) (see APPENDIX B for a sample of move-marked text). With

pairs of move tags positioned before and after text segments identified as certain moves,

for example, <IET2> and </IET2>, a concordance program could easily compute the

frequencies of the moves and steps in the texts. Most importantly, those move texts could

be grouped and saved as separate files for subsequent analyses of linguistic features.

3.2.3.3 Part of Speech Tagging

Subsequent to the move mark-up step, the article texts marked up with

move tags were split into move texts which were then grouped in separate folders,

resulting in a move corpus called the Education Research Corpus-Moves (ERC-M).

Part-of-speech   tagging was then performed to assign part-of-speech tags

to each word in the ERC_M. This type of annotation was necessary because untagged,

raw texts allow mainly word-based searches. The present study, which involved a

simultaneous analysis of a wide range of linguistic features in a fair-sized corpus,

depended heavily on category-based searches. That is, rather than searches for literal

words and word combinations in the corpus, the research questions to be answered by

an MD analysis required searches for lexical classes, grammatical categories, and syntactic constructions. Therefore, POS tagging became a prerequisite.

The ERC-M was automatically tagged by the Stanford POS Tagger (Toutanova et al., 2015). The tagger uses the Penn Treebank II tag set for the English language and is reputed for its token accuracy rate as high as 97.24% (Toutanova, Klein, Manning, & Singer, 2003) (see APPENDIX C for the Penn Treebank II Tag Set and APPENDIX D for a sample of POS tagged text).

3.2.4 Overview of the Education Research Corpus

The ERC contains 991,407 word tokens and 21,821 word types. The average length of the articles in the ERC is 8,131 words. The Introduction, Methods, Results and Discussion sections of the articles average 2,363 words, 1,844 words, 2,175 words, and 1,870 words, respectively.

Table 3.1 Overview of the ERC

Version	No. of Texts	Tagging Status
ERC-R	120	Raw
ERC-A	120	Move tagged
ERC-M	3905	POS and move tagged

As shown in Table 3.1, the ERC has three versions: a) ERC-R, which is a raw, untagged version with 120 full-length articles as text files; b) ERC-A, whose texts are marked up with labels of moves and steps identified in the corpus; and c) ERC-M, which is the result of splitting the articles in the ERC-A into moves and regrouping them into sub-folders of moves.

To facilitate data retrieval for different research objectives, text files are placed in folders which in turn are placed in sub-folders representing different tiers. The folders and files are named so that they enable intuitive recognition as well as easy manipulation when necessary.

The ERC-R has 12 folders, each representing a journal with its abbreviated 3-letter title as its name. Thus, for example, the folder "LAI" contains 10 articles from the journal *Learning and Instruction*. The 12 folders are AER, CAE, ESJ, HER, JEE, JLS, LAE, LAI, LLT, RST, SOE, and SSR. Inside each folder are ten article files. They are named thus: 3-letter abbreviated journal title_number of article. For example, SSR_01 refers to the first article from the journal *Scientific Studies of Reading*.

The ERC-A is identical with the ERC-R in terms of how folders and files are organized and named.

The ERC-M has 16 folders of move files. Each folder is named thus: initial letter of the section name _ 2-letter abbreviation of the move label. Thus, for example, the folder "IET" contains text segments identified as an *Establishing a territory* move in the introduction section. Inside the move folders are move text files which are thus named: 3-letter move code_3-letter journal code_number of article_number of move text. For example, the file IET_SSR_05_04 contains the fourth text of the *Establishing a territory* move identified in the introduction section of the fifth article taken from the journal *Scientific Studies of Reading*.

3.3 Move Analysis

As indicated in Section 3.1, corpus analyses in this study were conducted in two phases: analysis of the rhetorical structure and analysis of linguistic features in the texts. This section describes the first phase aimed at identifying the rhetorical moves and steps in each section of the 120 RAs in the ERC, determining their sequencing in each section, and based on the frequencies of moves and multi-move sequences generated, deriving a prototypical rhetorical structure of RAs in education research.

3.3.1 Analytical Framework

This study adopted Swales' move analysis (1981, 1990, 2004) as its analytical framework. Providing a comprehensive description of the overall rhetorical structure of education RAs was part of the ultimate goal of the study, which was to assist rhetorically and linguistically disadvantaged NNS education researchers in RA writing. The first objective of this study was to identify the rhetorical move structure of the articles in the ERC. As reviewed in Section 2.1.2, Swales' move analysis (1981, 1990, 2004) is particularly suited for revealing the rhetorical structure of genres in terms of communicative purpose. Within this framework, move analysis usually takes these three steps: 1) coding the rhetorical moves and steps of a set of texts representing the genre according to the local communicative purposes they realize, 2) determining which moves and steps are obligatory, conventional, optional or unconventional based on frequencies of occurrence, and 3) deriving a prototypical rhetorical structure for the genre based on the allowable order of the moves and steps.

3.3.2 Coding Scheme

An initial coding scheme was developed on the basis of previous studies on the research article move structure (see APPENDIX E). In particular, Swales' CARS model (1990) for the introduction section and Pho's (2008a) applied linguistics RA model made the most important contribution to the initial scheme.

While the CARS model has proved powerful in accounting for the move structure of RAs in many disciplines, Pho's 14-move model provides a detailed, all-four section description of the rhetorical structure of applied linguistics RAs. As discussed in Section 2.2.5, Pho's comparative study of applied linguistics and education RAs revealed similar move structures of RAs in the two related disciplines, both within the broader field of education. Applied linguistics RAs have fourteen moves, but all eleven moves with their component steps in education technology are found in applied linguistics.

The choice of Pho's model was motivated by its relevance to the general field of education research as well as its comprehensiveness in rhetorical structure characterization. The present study had a much wider scope covering such areas of education research as engineering education; management and business education; literacy education; education technology; health education; educational sociology; science education; applied linguistics; and learning theory and psychology. Yet, more commonalities than differences were expected between education RAs and RAs in applied linguistics and education technology, as both groups belong to the general field of education research.

3.3.3 Coding Process

3.3.3.1 Communicative-Purpose-Only Criterion

It was vital to have a clear, operational criterion for coding the texts for moves and steps. As the move is a functional, not a formal unit, this study assigned move tags to text segments depending on their content or communicative function only. Previous studies have used different criteria for move identification, including function-based (e.g., Kwan, 2006; Pho, 2008b; B. Zhang et al., 2012), form-based (e.g., K. Anderson & Maclean, 1997; Lim, 2006) and a combined approach of function and form (e.g., Kanoksilapatham, 2007; Nwogu, 1997; Swales, 1990). While the function-based approach can be subjective to a certain degree, the form-based approach that uses linguistic features as clues is fundamentally flawed as it is not in line with the concept of move (Paltridge, 1994). A rhetorical move, by definition, is a text segment that not only performs a specific communicative function of its own but also contributes to the overall communicative purpose of the genre. Linguistic features are only secondary to communicative purposes as they serve to realize the communicative purpose of the moves (Swales, 1981). The third approach of combining function and form has been criticized for its logical fallacy of circular reasoning by which linguistic features are used to identify or prove the existence of moves and then the moves are analyzed for linguistic features (Paltridge, 1994).

Identifying moves by the communicative purpose that the writer wishes to accomplish will inevitably result in varying move lengths. Some moves are realized by

one sentence only while others take more than one sentence or even one or more paragraphs to realize their communicative purposes. In this study, a move was operationally defined as a chunk of text of at least one complete sentence that fulfilled a distinct communicative purpose. Any sentence or group of sentences, regardless of length, was considered an instance of a move as long as it had a communicative purpose to realize.

3.3.3.2 Treatment of Move Embedding and Recycling

In cases of move embedding whereby more than one communicative purpose appeared to be realized within one sentence as described in Pho (2008b), following Holmes (1997), Ozturk (2007) and Hirano (2009), the sentence was identified as one single move by the most dominant communicative purpose. There were three reasons for disregarding move embedding. First, in contrast to genres of a compact nature like the RA abstract, the RAs in the ERC, with a minimum length of 5,000 words, were not expected to abound in embedded moves. Second, embedded moves would be too short to contain enough linguistic variation to be included in a factor analysis in the second phase of the study. And third, if embedded moves were to be identified within a sentence, in cases where the words expressing an embedded move constituted part of the main move, whether to count the words twice in calculating move lengths would become a difficult issue. The following excerpts illustrate this point:

1) Here again we tested for the interaction between embodied and objectified cultural capital but did not find an effect. ....(SOE_03)

2) One of the aims of the present study was to investigate whether phonological core deficits extend to Dutch dyslexic adults who, as yet, have not been studied. .... (SSR_03)

Excerpt 1 appears to serve two communicative purposes: describing the data analysis procedure and reporting a specific result. However, given that the sentence appears in the results section and that statistical analysis has been described earlier in the methods section, this sentence was identified as a *Reporting specific results* move.

Excerpt 2 was taken from one of the few articles where the *Establishing a niche* move was absent. This sentence was coded as a *Presenting the present research* move although *"Dutch dyslexic adults who, as yet, have not been studied"* does signify a gap in previous research. Contextual clues indicates that the primary purpose of this sentence is to introduce the reader to the present study.

In this study, different steps of the same move were frequently found to cluster together without being interrupted by a different move or a sub-heading. As such, they were considered one move, and when splitting the articles into move texts, they were treated as one text.

Move recycling has been observed in all the sections of RAs by many researchers (e.g., Amnuai & Wannaruk, 2013a; Peacock, 2002; Posteguillo, 1999; Stoller & Robinson, 2013; Yang & Allison, 2003). In this study, if a certain move

recurred after an intervening one in a cyclical fashion, each occurrence of that move was considered an individual instance of the move.

3.3.3.3 Hand-Coding Articles

The coding process did not begin until after a set of codes had been created to represent the moves and steps in the initial coding scheme. As a rule, a move code consists of three upper-case letters, the first indicating the section where it appears and the other two the move. For example, the code IET stands for an *Establishing a territory* move of the introduction and MRD for a *Presenting the research design* move of the methods section. In cases where a move was realized by steps or sub-moves, a numeral was added to the stem. Thus IET2 means Step 1 of an IET move. If a step had sub-steps, the sub-step is indicated by an upper-case letter. This exclusively applies to IEN1, the only step with sub-steps. For example, IEN1A and IEN1B refer to Sub-steps A and B of Step 1 of an IEN move, respectively. The move codes were intuitively understandable and therefore reduced much cognitive burden on the coders. In the actual coding process, the codes were hand-written in the margins of the articles to mark text segments for moves and steps. It should be noted that during the course of coding, some new codes were added to accommodate new moves and steps that emerged from the data.

The coding process began, once an operational definition of a move, an initial coding scheme and a set of codes were in place. It was a long and iterative process. The initial coding scheme served as a good point of departure in the analysis,

but throughout the process, new categories of functional units emerged and modifications to the coding scheme were made accordingly. As a result, many texts were coded and re-coded several times. Each revision of the coding scheme was only a closer approximation of an accurate description of the rhetorical structure of the RAs in the corpus. Yet the scheme was always open for fine-tuning until all the articles had been coded before it finally encompassed all the functional units eligible for move status. As shown in Chapter 4, a total number of 16 moves were identified in the 120 education RAs in the ERC.

3.3.3.4 Enhancing Reliability

Move analysis is primarily qualitative and therefore, special effort is necessary to ensure reliability. Neuendorf (2002) states that when human coders are used, reliability translates into inter-coder reliability. Inter-coder reliability has become a critical component of move analysis where it refers to the extent of agreement between each independent coder's assignment of all semantic functional units in the texts to moves and steps. Since Crooks (1986), many move analysts (e.g., Kanoksilapatham, 2007; Peacock, 2002; Pho, 2008a) have used independent coders to ensure a high degree of agreement in their analyses.

In the present study, an experienced internationally published education researcher was enlisted as inter-coder. She underwent extensive hands-on training by the researcher. First, the researcher explained to her the notion of move analysis,

guidelines for move identification, and the coding scheme. The coding scheme used at this stage had been calibrated with the addition of two new moves as a result of a pilot analysis of 24 articles (20% of the ERC corpus, 2 from each journal) by the researcher. It was also augmented with rich examples extracted from the analyzed articles.

Subsequent to the orientation phase, the researcher and the inter-coder together coded three texts randomly drawn from the corpus, during which course the researcher continuously monitored inter-coder reliability. The process acquainted the inter-coder with the methods of move coding.

Then, the researcher and the inter-coder each independently coded 36 texts (30% of the corpus) randomly drawn from the corpus. Inter-coder reliability was assessed for this batch of texts by comparing the researcher's and the inter-coder's results. The percentages of agreement on move boundaries and labels assigned for the introduction, methods, results and discussion sections were 85.6%, 92%, 89.4%, and 90.2%, respectively, averaging 89.3%. It is believed that such a satisfactory level of inter-coder reliability was to a large extent due to the enhanced coding scheme that was sensitive to the education RAs in the corpus. Additionally, the straightforward codes and rich illustrative examples contributed to inter-coder reliability as they posed little challenge for interpretation.

To resolve the coding discrepancies, discussions were held between the researcher and the inter-coder until 100% agreement was reached. The discussions proved very valuable for the researcher in the later stage of coding the remainder of

the corpus alone. Some differences turned out to be mistakes on the researcher's part due to insufficient knowledge of the education research field. By working with an experienced education researcher on a large portion of the ERC corpus, the researcher gleaned much knowledge of the discourse community of education researchers—their communicative purposes, stylistic conventions, and the complexity involved in their choice of strategies and linguistic expression. This enhanced knowledge considerably led to better decisions in coding the remaining 84 articles—the piloted texts were analyzed anew, hence improving the reliability of results.

In addition, a NS researcher was occasionally consulted about a few chunks of text whose communicative purposes seemed dubious.

3.3.4 Deriving the Prototypical Move Structure

The move coding process answered Research Question 1 by identifying all moves and their constituent steps used in the ERC RAs. It also led to the creation of the ERC-A and ERC-M required for use in addressing Research Question 2 as well as the other two research questions for the second phase of the study.

3.3.4.1 Obtaining Frequencies of Moves and Steps

Essentially, Research Question 2 concerns what moves and steps must occur and where; and what moves and steps may occur and where. The answer to this question is embodied in a prototypical generic structure of moves and steps classified according to their frequency of use and organized in their most likely order.

Most previous studies adopted a dichotomous criterion labeling moves and steps either as obligatory or optional based on an arbitrary cut-off percentage of the total number of RAs under study. For example, while a cut-off percentage of 50% was used in Nwogu (1997) and Loi & Evans (2010), a higher percentage of 60% was used in Kanoksilapatham (2005, 2011). Nevertheless, to increase descriptive accuracy for pedagogical purposes, this study classifies moves and steps into four status categories: obligatory ($x = 100\%$), conventional ($60\% \leq x < 100\%$), optional ($5\% \leq x < 60\%$), and unconventional ($x < 5\%$).

Technically, in order to answer Research Question 2, frequencies of each move and its distribution over the entire corpus had to be obtained and its position in relation to other moves had to be determined. To this end, additional .txt files and folders were prepared. First, a file of move and step codes was prepared using EditPat Light (Goyvaerts, 2015) which extracted the codes from the files in the ERC-A and merged the files into one. This file contained exclusively all the move and step codes in the order as they appear in the articles. Then, to produce files of move codes only, the file of move and step codes was edited by deleting the suffixes of the step codes (i.e., numerals and capital letters for steps and sub-steps) and then, in cases of any code recurring more than once without being intervened, deleting the redundant ones. Next, this master move codes file was split into 120 files, one for each article in the ERC corpus. Finally, the move codes in the 120 move codes files were grouped by

sections and saved in four separate folders named by their section names. Thus, for example, the folder named move_sequence_introduction refers to the folder containing 120 files of the original moves in the introduction sections. These folders are called section move sequence files hereafter.

Frequencies of moves and steps, which provided an empirical basis for determining which moves and steps are obligatory, conventional, optional or unconventional were obtained by running the move-tagged ERC-M corpus through the concordance function of AntConc (Anthony, 2014).

3.3.4.2 Determining the Order of Moves and Steps

To decide the typical order in which moves tend to appear in each section of the education RAs, following Feng (2006) and Chang and Kuo (2011), this study resorted to examining multi-move sequences. The section move sequence folders provided data input for this purpose. In using AntConc (Anthony, 2014) to extract the most salient multi-move sequences that are widely distributed in the articles, both the minimum frequency and minimum range were set at 35.

At this juncture, it was tempting to rely on move collocations to determine what other moves a move was strongly associated with. In phraseology studies, collocations are two words that appear in tandem more frequently than their individual frequencies would suggest (Hoey, 1991; Sinclair, 1987, 1991; Stubbs, 1995). Collocates predict each other in that where one item appears, chances of finding the other

increase. An analogy to lexical collocations, move collocations are pairs of moves that have a propensity for co-occurrence. Theoretically, on the basis of move collocations, it is possible to derive high frequency cyclic move patterns and a prototypical move structure for each RA section. In practice, however, it would be difficult to interpret the massive amount of information on multiple lists of move collocates.

By comparison, information about multi-move sequences is much more useful and easier to handle. Multi-move sequences or move bundles are in fact extended move collocations containing two or more moves. A concordance program can simultaneously extract sequences of different sizes. This would perfectly fit the interest in how different numbers of moves cluster frequently in the corpus. Equally important, a concordance program produces a single list of multi-move sequences, which makes interpretation of results much easier than having to refer to multiple lists of collocates of a large number of moves.

3.4 Multidimensional Analysis

The second objective of this study was to map out linguistic co-occurrence patterns typically associated with the communicative purpose of each rhetorical move identified in the move analysis. Essentially, the task was to explore what linguistic features (observed variables) together are addressing the same underlying communicative functions (factors or constructs) in the rhetorical moves across the

education RAs. The assumption is that if linguistic features co-occur at a significantly high frequency, they are indicative of a shared underlying communicative function (Biber, 1988).

3.4.1 Analytical Framework

The present study employed multidimensional (MD) analysis (Biber, 1986, 1988) to simultaneously examine a large number of linguistic features used in the rhetorical moves of the 120 education research articles.

As reviewed in Chapter 2, previous linguistically-oriented studies have typically been concerned with a few opportunely selected linguistic features assumed to be discrete. The keyword approach as used in those studies of a few individual words or word combinations provided a less comprehensive contrast of text varieties and might not work for more fine-grained types of genre analysis involving a wide range of lexico-grammatical features (Xiao, 2009; Xiao & McEnery, 2005). To overcome the inherent weaknesses of the keyword approach, Biber and his followers (e.g., Biber et al., 2003; Kanoksilapatham, 2007; Xiao, 2009; Xiao & McEnery, 2005) have successfully investigated textual variations among language varieties and genres by examining how linguistic features co-occur to express basic communicative functions called dimensions. The purpose of MD analysis is to identify the functional dimensions of linguistic variation among texts and to provide an overall description of relations among groups of texts with respect to these dimensions. It is assumed that each of the

features has a functional basis. Features from the same category can have different functions, and features from different categories can have a shared function. This corpus-based approach has yielded more generalizable results about overall patterns of linguistic variation.

The core of MD analysis is the use of the statistical procedure of factor analysis in analyzing a large set of linguistic features. As noted by Pedhazur and Schmelkin (1991), factor analysis is "probably the most useful for studying the internal structure of a set of variables or indicators" (p. 66). Its power lies in its ability to render variable complexity into greater simplicity by reducing a large number of observed variables into a significantly smaller, more manageable set of factors of unobserved, underlying constructs (Kerlinger, 1979). These constructs, which are the very things the researcher wishes to understand, are theoretically understood as actually determining the observed scores on the measured variables (B. Thompson & Daniel, 1996). In essence, factor analysis is a method of categorizing variables which have something in common through identifying structures and commonalities in the relationships between observed variables.

3.4.2 Linguistic Features and Frequency Counts

This study investigated the same 67 linguistic features that Biber investigated in several of his MD studies (e.g., 1988, 2009) (see APPENDIX F for a list of the features). This set of features represent different lexical, grammatical, semantic and even

pragmatic categories (e.g., verbs, adjectives, pronouns, past tense, present perfect aspect, passive voice, relative clauses, adverbial clauses, complement clause types, public verbs, suasive verbs, hedges, downtoners). Biber's choice of these features was informed by a careful survey of previous research. "As such, these features provide a solid basis for determining the underlying functional dimensions in English" (1988, p. 72).

The MD analysis in this study depended on the frequency counts of the linguistic features in the 3,905 move texts in the ERC-M. Since the tags for RA tiles, headings, moves, foreign words, mathematical formulas, and excerpts were POS tagged as nouns, these irrelevant tags had been deleted in advance to avoid inflated frequency counts of nouns. The AntConc (Anthony, 2014) concordance program was used for obtaining frequencies of the linguistic features, which were then normalized to per 100 tokens so that they had a common basis for comparison.

Since some features had a frequency value of zero in many move texts, including them as they were in the analysis would have meant an inadequate number of cases, reducing the legitimacy of a factor analysis. It would have been even less appropriate to simply exclude them from analysis as this would have harmed data integrity and led to an incomplete picture of inter-move linguistic variation. Following Kanoksilapatham (2007), they were collapsed into super-ordinate categories as composite variables. For example, contractions (e.g., *n't*), subordinate *that* deletion (e.g.,…*said it was important*), stranded prepositions (e.g. *the book that I was thinking*

of), split infinitives (*to convincingly prove*), and split auxiliaries (e.g., *they are objectively shown that ...*) were rarely used in the ERC corpus because they are characteristic of speaking genres rather than written genres. Therefore, they were combined into one single feature named dispreferred forms. This treatment reduced the original set of variables to 36, still representing the original 67 features.

Before performing factor analysis, texts of less than 30 words, reflecting little linguistic variation, were excluded because, otherwise, they would have skewed the subsequent factor analysis aimed at capturing linguistic variation. At the same time, inflated type/token ratios were avoided too. Similarly, texts of the *Describing the conceptual model* and *Summarizing results* moves were also excluded due to their limited observations. This treatment reduced the dataset from 3,905 to 3,421 texts, containing 939,084 tokens as opposed to the original 991,407. With only a 5% reduction, the representativeness of the data was preserved.

Table 3.2 Descriptive Statistics of Linguistic Features

Linguistic features	Range	Minimum	Maximum	Mean	Std.
Type/token ratio	61.00	22.00	83.00	60.69	13.25
Nouns	46.54	6.00	52.54	25.27	5.24
Prepositional phrases	27.27	0.00	27.27	12.14	2.89
Attributive adjectives	26.09	0.00	26.09	9.52	3.61
Nominalizations & gerunds	30.23	0.00	30.23	6.57	3.27
Mean word length	3.93	3.41	7.34	5.53	0.45
Present tense	16.22	0.00	16.22	3.83	2.65
Adverbs	21.88	0.00	21.88	3.02	1.94
Past tense	14.04	0.00	14.04	2.86	2.59
Personal pronouns	12.28	0.00	12.28	1.85	1.80
Private verbs	11.11	0.00	11.11	1.64	1.41
to-infinitives	10.81	0.00	10.81	1.58	1.32
Demonstratives	9.38	0.00	9.38	1.52	1.28
Passives	9.09	0.00	9.09	1.39	1.29
Phrasal coordination	12.12	0.00	12.12	1.31	1.25
Be as main verb	9.26	0.00	9.26	1.25	1.15
Participial clauses	8.82	0.00	8.82	1.00	1.03
Modals	9.38	0.00	9.38	0.92	1.19
Relative clauses	6.66	0.00	6.66	0.86	0.90
Pragmatic expressions	8.57	0.00	8.57	0.86	1.00
Independent clause coordination	6.82	0.00	6.82	0.78	0.84
Conjuncts	10	0.00	10.00	0.77	0.90
Predicative adjectives	8.11	0.00	8.11	0.74	0.90
Public verbs	8.86	0.00	8.86	0.63	0.89
Complements	9.38	0.00	9.38	0.60	0.79
Negations	9.38	0.00	9.38	0.59	0.83
Dispreferred forms	12.5	0.00	12.50	0.56	0.82
Adverbial clauses	6.25	0.00	6.25	0.52	0.72
Suasive verbs	6.98	0.00	6.98	0.45	0.68
Present perfect aspect	7.69	0.00	7.69	0.39	0.75
Pronoun *it*	8.16	0.00	8.16	0.38	0.67
Existential there	5.41	0.00	5.41	0.16	0.42
Seem/appear	5.38	0.00	5.38	0.11	0.35
Pro-verb do	3.45	0.00	3.45	0.04	0.20
Direct Wh-questions	5.36	0.00	5.36	0.02	0.19
Independent pronouns	2.56	0.00	2.56	0.01	0.10

Table 3.2 displays descriptive statistics of the 36 linguistic features in the remaining 3,421 move texts in the descending order of means. The frequencies have been normalized to per 100 tokens. The mean scores range from 0.01 to 60.69,

indicating great differences in their use. Similarly, the standard deviations range from

0.10 to 13.25, reflecting different amounts of variation within the features.

3.4.3 Factor Extraction

Specifically, to identify the co-occurring linguistic features, the present study,

following previous researchers (Getkham, 2010; Kanoksilapatham, 2007), invoked

Principal Component Analysis (PCA). The dataset of the normalized frequencies of the

36 variables was analyzed with the SPSS statistical package (v. 16.0.2) (IBM, 2008).

PCA reduces the number of a large set of observed variables to a smaller set of

variables that have strong linear inter-correlations. These new variables are called

principal components which can explain a large portion of the total variance of the

original observed variables. The number of components extracted is equal to the

number of observed variables in the analysis. Eigenvalues are indicative of the

amount of variance explained by each component. The first few components with

greater Eigenvalues are retained as principal components, which account for

progressively less of the variance in the data. Together, they explain a maximal

amount of variance of the observed variables.

To assess the suitability of the data for a PCA, three indicators were examined.

The KMO test showed that the MSA (Measurement of Sampling Adequacy) value of

0.43 was a bit low. However, Bartlett's Test of Sphericity rendered an associated *p*

value of 0.001, and the summary measure of the correlation matrix, the determinant

value, was 0.002, both suggesting that the variables were significantly correlated. In view of the relative textual homogeneity of the education research RA genre, the MSA value of 0.43 was not an issue. Therefore, the dataset was considered meeting the assumptions for a valid PCA.

To determine an optimal number of components to retain, three procedures were undertaken. First the Eigenvalues were inspected. A total of 14 linguistic features had initial Eigenvalues greater than 1. Kaiser's (1960) Eigenvalues-greater -than-1 criterion suggested a 14-component model, a number too large to enable efficient interpretation. In fact, as pointed out by Leandre et al. (1999), this criterion is frequently unreliable because it tends to result in substantial over-factoring. As recommended by Cattell (1966), the scree plot of the Eigenvalues of the correlation matrix was subsequently examined (see Figure 3.2). The graph showed the last substantial drop in the magnitude of the Eigenvalues was at Component 7. According to this procedure, a model with the same number of principal components as the number of Eigenvalues before the drop would then fit to the data. That would mean a 6-component model. Next, two separate PCAs were performed to try a 6- and 8-component solution to see how many linguistic features saliently loaded on the components. In the 6-component solution, salient loadings ranged from five to eleven, but in the 8-component solution, one component had only two salient loadings. According to Biber (1988), a minimal of five salient loadings were required for a

meaningful interpretation. Thus, a 7-component model was selected as it represented

a balance between completeness and parsimony, i.e., an optimal amount of the total

variability explained and the least number of components to explain it.
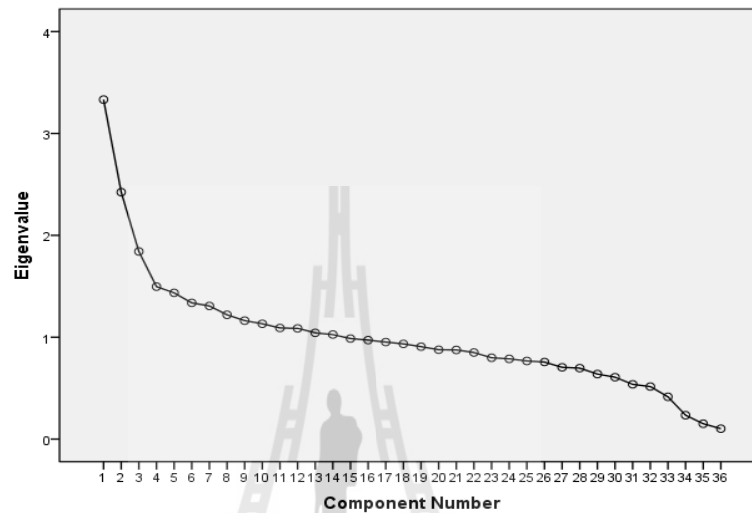
Figure 3.2 Scree Plot of Eigenvalues of Linguistic Features

Therefore, a 7-component PCA was performed on the data. Following Biber

(1988), Promax rotation was used to allow the components to correlate with one

another. The results of this analysis will be reported in Chapter 5.

Table 3.3 Variance Explained by the Seven Factors

Component	Initial Eigenvalues			Extraction Sums of Squared		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.3	9.3	9.3	3.3	9.3	9.3
2	2.4	6.7	16.0	2.4	6.7	16.0
3	1.8	5.1	21.1	1.8	5.1	21.1
4	1.5	4.1	25.3	1.5	4.1	25.3
5	1.4	4.0	29.3	1.4	4.0	29.3
6	1.3	3.7	33.0	1.3	3.7	33.3
7	1.3	3.6	36.6	1.3	3.6	36.6

Table 3.3 displays the percentage of the overall variance explained by each individual component and all the seven principal components together. As can be seen, the seven principal components account for 36.6 % of the overall variance, an amount considered satisfactory given the relative homogeneity of the genre of education research articles (Kanoksilapatham, 2007) (see APPENDIX G for total variance explained by all the components)

3.4.4 Characterizing Inter-move Variation along the Dimensions

This procedure was aimed at specifying the similarities and differences among the moves with respect to each of the dimensions of linguistic variation identified in the factor analysis. The factor score of a text with respect to a certain factor indicates the location of the text on that factor. The extent to which moves are similar or different in the use of clusters of linguistic features is revealed when the mean factor scores of each of the moves are plotted along each of the dimensions.

The present study utilized regression factor scores (Thurstone, as cited in DiStefano, Zhu, & Mîndrilă, 2009) obtained with SPSS to characterize moves with respect to the identified dimensions. Regression-based factor scores are z-scores with a mean of zero and a standard deviation of 1 of the distribution of factor scores if PCA is used as in the present study. The regression approach to estimating factor scores is known as a refined method because it takes into account not merely the correlation between factors extracted and that between the factors and manifest variables (in this

case, linguistic features), but also the correlation among observed variables and that among oblique factors. According to DiStefano, Zhu, and Mîndrilă (2009), this process maximizes the validity of factor scores which are thus advantageous over factor scores expressed by the summed standardized frequencies of the linguistic features as used in previous MD analyses of linguistic variation (e.g., Biber, 1986; Biber, 1988; Cao & Xiao, 2013; Kanoksilapatham, 2007; Xiao, 2009).

3.4.5 Testing for Inter-move Differences along the Dimensions

To find out whether the moves were significantly different with respect to each dimension, seven ANOVAs were performed. The results of all seven homogeneity tests were significant, $p < 0.5$, indicating significant differences among the moves with respect to all seven dimensions. This also meant that the assumption of homogeneity of variance was violated and that certain procedures were required to rectify the matter so that pairwise comparisons could be made to determine whether there was a significant difference between each possible pair of moves. However, this matter was further complicated by unequal sample sizes (number of texts of each move sub-corpus). In the ERC, the number of texts of each move varied wildly from 54 to 603. With unequal sample sizes, ANOVA is not robust to violations of homogeneity of variance. Consequently, no *post hoc* procedure was taken to make pairwise comparisons. Instead, mean dimension scores were used as a basis for determining differences between moves.

3.5 Summary

This chapter provided detailed descriptions of the research design and the procedures of data collection and analysis. The research design is characterized by the integration of a qualitative move analysis and a quantitative MD analysis of linguistic variation. Specifically, detailed information was provided on corpus compilation and the two constituent parts of corpus analysis, namely, the move analysis and MD analysis. Meanwhile, justification was given for every important decision on such matters as research design, journal and article selection, corpus annotation, analytical frameworks, and data analysis procedures.

Chapter 4 will present and discuss the results of the move analysis in answer to Research Questions 1and 2.

CHAPTER 4

RESULTS AND DISCUSSION I:

MOVES AND GENERIC MOVE STRUCTURE

This chapter reports and discusses the results of the move analysis conducted on the ERC corpus. Section 4.1 presents the results in five subsections. Subsection 4.1.1 is an overview of the moves and their distributions in the different RA sections. Subsections 4.1.2-4.1.5 report the results with respect to each of the four sections of the RAs in the IMRD order. In Section 4.2, the results are discussed in subsections 4.2.1-4.2.5, with respect to the four RA sections.

4.1 Results

The first phase of the study was aimed at analyzing the rhetorical move structure of education RAs. As expounded in Section 3.1, two research questions were raised for this phase: 1) What are the rhetorical moves and constituent steps employed in each section of education RAs as represented by the ERC? and 2) How are the moves and constituent steps typically sequenced in each section of education RAs as represented by the ERC? To address these questions, Swales' (1981, 1990, 2004) move analysis framework and an initial coding scheme incorporating findings of several previous studies were applied to the data, i.e., the ERC corpus. The process

was basically qualitative in that the identification of moves and demarcation of their boundaries involved human judgment. Meanwhile, frequency-based quantitative analysis played a key role in disclosing the overall rhetorical structure of the ERC articles.

In the subsections that follow, first an overview is presented of all the rhetorical moves identified and their distributions in the four IMRD sections of the articles. Then for each RA section, a sub- subsection presents findings about each move and step observed, and a prototypical move structure of the RA section.

4.1.1 Overview of Moves and Steps in Education RAs

Tables 4.1, 4.3, 4.5 and 4.7 display all the moves and steps identified in the ERC with their descriptive statistics and status. Note again that the moves are mnemonically labeled with three capital letters to assist memory in the coding process and for facility in corpus searches. The first letter is the initial of the section heading where the move belongs and the next two abbreviate the name of the move. Therefore, for example, "IET" refers to *Establishing a territory* of the introduction section while "MDC" stands for *Describing data collection procedures* of the methods section.

Overall, 16 moves with their constituent steps have been identified, of which 5 are obligatory, 7 conventional, and 4 optional. By comparison, the move structures of the sections broadly resemble the corresponding sections of the a priori coding scheme, indicating that the scheme was satisfactorily comprehensive and sensitive to

the textual realities of the ERC articles. Nevertheless, differences do exist mostly due to additions of new moves and steps for a more faithful and comprehensive delineation of the ERC articles under investigation.

New moves and steps have been added to categorize text segments with distinct communicative purposes of their own that were not captured by the initial coding scheme. Specifically, additions to the introduction section include a *Describing the conceptual model (IDM)* move with its three constituent steps (IDM1, IDM2 and IDM3); a *Presenting a vignette (IET1)* step and a *Defining terms and concepts (IET5)* step to the *Establishing a territory (IET)* move; a *Providing positive justifications (IEN2)* step to the *Establishing a niche (IEN)* move and a *Postulating hypotheses (IEN1E)* sub-step to IEN Step1; and *Claiming significance of the research (IPR4)*, *Verifying compliance with ethical standards (IPR5)* and *Describing research context (IPR6)* steps to the *Presenting the present research (IPR)* move.

To the methods section, additions include a *Describing research design (MRD)* move; a *Describing research context (MDC1)* step and a *Verifying compliance with ethical standards (MDC5)* step to the *Describing data collection (MDC)* move; and a *Recounting data analysis procedures (MDA1)* step, a *Justifying data analysis procedures (MDA2)* step and an *Establishing inter-coder reliability (MDA3)* step to the *Describing data analysis procedures (MDA)* move.

The results section has a smaller number of additions, including a *Summarizing results (RSR)* move, and three steps added to the *Reporting specific results (RRR)* move: *Preparing for specific results presentation (RRR1)*, *Presenting specific results (RRR2)*, and *Illustrating with examples or excerpts (RRR3)*.

As for the discussion section, a *Summarizing the study (DSS)* move has been added. At the step level, a *Restating hypotheses or research questions (DCR1)* step, a *Restating results (DCR3)*, an *Illustrating with examples or excerpts (DCR6)* and a *Recommending future research (DCR7)* step have been added to the *Commenting on specific results (DCR)* move. Similarly, the *Evaluating the study (DES)* move has an additional *Indicating significance (DES1)* step.

Apart from additions of moves and steps, the tables also reflect some deletions, merges and renamed moves and steps. The *Drawing conclusions of the study* and *Highlighting overall findings of the study* moves have been removed as their communicative functions are well covered by the newly added *Summarizing the study (DSS)* move. The *Discussing the findings of the study* move has been renamed *Commenting on specific results (DCR)* as the original label means virtually the same as that of the entire section. Additionally, for consistency, the label *Deductions from the research* has been replaced with the gerund phrase *Making deductions from the research.* Also note that due to the additions, the steps have been re-ordered.

4.1.2 The Introduction Section

Table 4.1 below displays the four moves observed in the introduction sections of the articles: *Establishing a territory (IET)*, *Establishing a niche (IEN)*, *Describing the conceptual model (IDM)*, and *Presenting the present research (IPR)*, all with a number of constituents steps. While IET and IPR were obligatory and IEN highly conventional, IDM was optional. Details of the findings are presented in subsequent sub- subsections.

Table 4.1 Rhetorical Move Structure of the Introduction Section

Code	Move	f	n	%	S
IET	**Establishing a territory**	**1318**	**120**	**100**	**Obl**
IET1	Presenting a vignette	5	5	4.2	Unc
IET2	Claiming centrality	114	75	62.5	Con
IET3	Making topic generalizations	535	116	96.7	Con
IET4	Reviewing items of previous research	582	117	97.5	Con
IET5	Defining terms and concepts	82	54	45	Opt
IEN	**Establishing a niche**	**411**	**114**	**95**	**Con**
IEN1	(one or more of A-E)	358	111	92.5	Con
IEN1A	Counter-claiming	69	46	38.3	Opt
IEN1B	Indicating a gap	203	91	75.8	Con
IEN1C	Raising a question	49	37	30.8	Opt
IEN1D	Continuing a tradition	30	25	20.8	Opt
IEN1E	Postulating hypotheses	7	4	3.3	Unc
IEN2	Providing positive justifications	53	41	34.2	Opt
IDM	**Describing the conceptual model**	**52**	**21**	**17.5**	**Opt**
IDM1	Introducing the model	20	20	16.7	Opt
IDM2	Identifying components	8	8	6.7	Opt
IDM3	Describing components' relationships	24	18	15	Opt
IPR	**Presenting the present research**	**497**	**120**	**100**	**Obl**
IPR1	Announcing the present research	392	119	99.2	Con
IPR2	Announcing principle findings	8	8	6.7	Opt
IPR3	Indicating RA structure	43	37	30.8	Opt
IPR4	Claiming significance of the research	34	27	22.5	Opt
IPR5	Verifying compliance with ethical standards	1	1	0.8	Unc
IPR6	Describing research context	19	17	14.2	Opt

Note. f = frequency of a move/step in all the RAs; n = number of RAs in which a move/step appears; % = percentage of a move/step in all the RAs; S = Status; Obl = obligatory; Con = conventional; Opt = optional; Unc = unconventional

4.1.2.1 Establishing a Territory

As Table 4.1 shows, the first move *Establishing a territory (IET)* was by far the most frequent in the introduction sections (f = 1318) and was observed in all the RAs (n =120). All but one introduction text opened with this move. By this obligatory opening move, the authors introduced the reader into a relatively general field of research to which the study reported thereafter was intended to make a contribution. This communicative purpose was expressed by varied combinations of the following steps or sub-moves.

4.1.2.1.1 Presenting a Vignette

A *Presenting a vignette* step *(IET1)* was found at the onset of a few introductions. As Table 4.1 shows, this step deserves only an "unconventional" label (4.2%; f = 5; n = 5). Nevertheless, this finding provides evidence that occasionally, in order to provide insight into a particular situation or topic, a writer can pose a real or hypothetical scenario of a certain activity or situation.

> *1) Years ago, while at a practice session at the U.S. Ballroom Dancesport Championships, I observed a tall, sort of gangly-looking fellow with headphones on dancing around to his music by himself. I thought that was weird, and figured that he was not a very good dancer. I commented about this to a friend who said, "Oh, that's Hans Laxholm, the runner-up world champion!" Suddenly, I watched him intently with a new eagerness to learn as I saw how incredible he was…(LAE_06)*

Excerpt 1 is a brief sketch of the author's personal experience. By sharing it with the reader, the author brings up the general topic of how false perceptions based on inadequate information can lead to a wrong judgment. Indeed,

the author immediately starts to review previous research on this very topic.

2) Consider two PhD students pursuing degrees in management. The first is affiliated with a well-established doctoral program that is known for the quality of its faculty members' scholarly output and their leadership in professional organizations. In contrast, the second student is enrolled in a less established program, with fewer research-oriented and professionally involved faculty. Both students, however, have authored papers in top-tier management journals. They have similar teaching portfolios and evaluations, requisite conference presentations, and nearly identical involvement in professional activities, including serving as paper reviewers for conferences and as discussants at regional and national Academy of Management meetings. ....(LAE_01)

Excerpt 2 juxtaposes two persons with similar scholarly records but different affiliations, one enrolled in a well-established PhD program and one in a less established one. Immediately after this short sketch of two imaginary scholars, the author raises questions regarding the impact their different affiliations would have on their career opportunities, which constitutes an *Establishing a niche (IEN)* move (see 4.1.2.2 below). In so doing, the author creates an interesting effect on the reader.

It is noteworthy that all five IET1s found in the corpus were restricted to two journals. One was *Academy of Management: Learning and Education* specializing in management education and the other *Journal of Learning Sciences (JLS)* in the field of learning sciences. Perhaps, this strategy of territory establishment is journal-specific.

4.1.2.1.2 Claiming Centrality

A conventional step found in 62.5% (n = 75) of the articles was *Claiming centrality* (IET2) which appeared once or twice in each article (see Table

4.1). In making a centrality claim, a majority of the writers made "appeals to the discourse community whereby members are asked to accept that the research about to be reported is part of a lively, significant or well-established research area" (Swales, 1990, p. 144). In conformity with Swales' observation, centrality claims opened introductions of the ERC articles typically in a single sentence. As exemplified below, centrality claims were frequently found to be confident assertions about the importance of a research topic or statements about research activities and interests.

3) Documenting the preparation of students for university-level mathematics as a function of their high school mathematics curriculum is essential to remedying the problem of underprepared students...(AER_02)

4) In the past fifty years or so, much research on the composing process has endeavored to understand the workings of the human mind as it constructs a text. ....(LLT_09)

4.1.2.1.3 Making Topic Generalizations

Making topic generalizations (IET3) was a highly conventional step observed in all but four RAs in the corpus (96.7%; f = 535; n = 116), with an average of five instances in each article (see Table 4.1). This step established a territory with statements about knowledge, practice, or phenomena, frequently in combination with centrality claims (IET2). In the ERC, this step was more frequently employed to launch an introduction. Unlike centrality claims, topic generalizations were found to vary in length from a single sentence to whole-length paragraphs.

In the introduction texts, topic generalizations were usually realized with statements of knowledge or practice with regards to the research topic as

shown in Excerpts 5 and 6, or as Excerpt 7 illustrates, by describing a relevant

phenomenon worthy of investigation.

> *5) Recent questions have been raised as to the effectiveness and relevancy of business education in general, and MBA programs in particular, about their ability to infuse students with the skills needed to function at an executive level in modern organizations ....(LAE_05)*

> *6) Over the past 25 years, researchers have employed numerous approaches to determine the relative value of private school attendance.........(ESJ_04)*

> *7) The underrepresentation of women in MBA programs is striking, particularly in comparison with medical and law schools in the United States, where numbers are much more balanced, (Catalyst, 2000). ....(LAE_02)*

IET3s in the corpus were often marked by a gradual progression from the most

to the least general to position the present research in an increasingly narrower

context. In other words, IET3s in the corpus typically started with bold

generalizations and ended with specific statements marked as tentative.

4.1.2.1.4 Reviewing Items of Previous Research

Reviewing items of previous research (IET4) was a very important

conventional step observed in the ERC in terms of the total number of occurrences (f =

582) and distribution (97.5%; n = 117) (see Table 4.1). By describing and critiquing what

had been found or claimed in past research, the absolute majority of the writers

established a space for their own work. IET4s usually dealt with specific studies in detail

as in Excerpts 8 and 9 or they might summarize groups of studies as in Excerpt 10.

> *8) In an influential study, Treiman and Zukowski (1991) demonstrated that, among English-speaking children, phonological awareness emerges gradually such that by the age of 4 years children can reliably make judgments about the similarity of syllables and by the age of 5 years such judgments*

can be made about subsyllabic units—namely, onsets and rhymes—but typically it is not until children begin formal schooling (age 6-7 years) that they are able to make reliable judgements about phonemes...(SSR_09)

9) On the other hand, recently, Hoffler and Leutner (2007) revealed in a meta-analysis of 26 studies on non-interactive animations that animations prove advantageous, especially when moderator variables like realism, depicted topic, and type of knowledge to be learned are controlled. ....(LAI_08)

10) Empirical studies have shown that concerns about habits of mind are more than esoteric debates. For example, students' epistemological beliefs are correlated with their learning approaches (Rodriquez & Cano, 2006; Zhu, Valcke, & Schellens, 2008), meta-cognitive learning preferences (Chuang & Tsai, 2005), and ability to monitor task difficulty and to calibrate their goal setting and planning (Stahl, Pieschl, & Bromme, 2006). ....(CAE_08)

As the excerpts show, citations were always provided to acknowledge works cited. Although they literally dotted all the four sections of the articles, citations, either integral as in Excerpts 8 and 9 or non-integral as in Excerpt 10 (see Swales, 1990, p. 114), had a much higher concentration in introductions, many of which had a "Literature Review" sub-heading or content-specific sub-headings. This finding points to the usefulness of IET4s in establishing a research space and validating a study by relating it to relevant previous research.

4.1.2.1.5 Defining Terms and Concepts

Table 4.1 shows a *Defining terms and concepts (IET5)* step observed in nearly half of the RAs in the ERC (45%; n = 54), hence an optional step. As illustrated by the excerpts below, definitions of terms were occasionally provided to clarify concepts or operationalize variables.

11) We use the term schema to refer to a generalized description of a word-problem type that requires similar solution methods (Citation)...........(ESJ_05)

12) For our study, physical environment includes availability and accessibility (AA) of FV, while social environments include parental factors associated with FV intake. .... (HER_01).

4.1.2.2 Establishing a Niche

The *Establishing a niche (IEN)* move was found to be a conventional move with a very high frequency and wide distribution across the corpus (95%; f = 411; n = 114). Typically, after territory establishment, the authors made this move to identify a specific space where the present research could occupy, although four out of 120 introductions in the corpus began by establishing a niche. It was also found that six introduction texts lacked an *Establishing a niche (IEN)* move. As the excerpt below illustrates, such rare cases is explained by an IEN move imbedded in a *Presenting the present research (IPR)* move (see 4.1.2.4).

13) One of the aims of the present study was to investigate whether phonological core deficits extend to Dutch dyslexic adults who, as yet, have not been studied. ....(SSR_03)

Specifically, the communicative purpose of niche establishment was fulfilled in one or a combination of the following steps and sub-steps.

4.1.2.2.1 Step 1 of Establishing a Niche

Table 4.1 shows that *Step 1 of Establishing a niche (IEN1)* was employed by the vast majority of the authors (92.5%; f = 358; n = 111).

Step 1 was occasionally realized by a *Counter claiming (IEN1A)* sub-step, stating that something was wrong with previous research, as the excerpts below demonstrate.

14) Although the NAEP dataset has many advantages, in particular its size, it includes a very limited number of control variables so selection bias is a significant concern for any studies using these data..........................(ESJ_04)

15) The description of learning activities was not focused on teachers' thinking processes associated with their visible activities. Different teachers who are apparently engaged in the same visible activity may actually use quite different thinking processes leading to different learning outcomes. ....(LAI_10)

Clearly, *Indicating a gap (IEN1B)* was the most dominant sub-step

(75.8%; f = 203; n = 91), by which writers pinpointed certain limitations in previous

research.

16) However, no investigator has examined the effects of linguistic statistical regularities on spoken and written word learning concurrently. .....(SSR_08)

17) There is little evidence of negative effects beyond these studies, though, and on current data, such outcomes seem to be neither widespread nor consistent...(LAI_03)

The above excerpts illustrate an important way of indicating a

niche, namely, by negating a certain aspect of previous research.

Question-raising (IEN1C) seems to be a minor niche

establishment strategy (30.8%; f = 49; n = 37). As the excerpts below show, direct

(with a question mark) or indirect questions (without a question mark) might be

asked.

18) But is it sound to assume that the causal mechanisms suggested by the ELM to account for the effect of information about the perceived competence of an information source on attitude formation can also account for the impact of quality information on media-based learning outcomes?... (LAI_09)

19) The question that logically follows is whether intellectual levels influence NOS teaching practice at the early childhood level.......................... (RST_01)

Another unimportant sub-step was *Continuing a tradition (IEN1D)* (20.8%; f = 30; n = 25). It was employed to express a need to approach a problem in a different way or to follow in the footsteps of previous researchers in solving a remaining problem.

> *20) A new answer may reside in constructing an additional layer of data that reflects, and thus, allows the study of, the cognitive processes underlying the construction of text...(LLT_09)*

> *21) While these studies provide useful information on learners' processing of captioned video, more studies are needed with a variety of learners and viewing conditions. ....(LLT_06)*

Although quite rare (3.3%; f = 7; n = 4) and thus considered unconventional, a *Postulating hypotheses (IET2E)* sub-step was identified in a few introductions.

> *22) Thus, same-gender students are probably seen as being more similar in educational achievements and therefore as providing more informative social comparison standards. This means that we can expect that students' academic self-concept will be affected more strongly by the achievement of gender ingroup than of gender outgroup classmates. ....(SOE_10)*

> *23) There should be substantial differences in how the capabilities of these two classes of video tools contribute to the context of collaborative learning through design. In principle, both types of tools can be used in design problems. However, collaborative video editing tools, although they may be more complex and demanding, might better support meaningful learning than video playback tools with word processors. .... (JLS_07)*

As illustrated by the excerpts above, the writers occasionally made hypothetical statements in establishing a niche. However, none of the writers used the term *hypothesis* when doing so. Instead, they used such expressions as "we can expect that..., " "There should be …differences in…," and "We argue that…" to make a prediction.

4.1.2.2.2 Providing Positive Justifications

Providing positive justifications (IEN2) was found to be an optional step present in one third of the RAs under investigation (34.2%; f = 53; n = 41) (see Table 4.1). It was much more frequently used in combination with IEN1B than on its own. After indicating a gap, the writers sometimes gave positive reasons why the present study was necessary.

> *24) Given the nascent stage of research in this area, empirical elaborations of central office administrators' participation can help future researchers better determine whether learning outcomes stem from the design of the initiatives or from how central office administrators participate in their implementation..(AER_04)*

> *25) It could therefore be useful to consider the types of speech and language difficulties shown by these children and whether any difficulties in particular are more closely linked to literacy difficulties................................(SSR_04)*

In Excerpt 24, the author suggests an empirical approach to the same question, saying it may help later researchers "better" identify the cause of effective learning. Likewise, the author of Excerpt 25 asserts that it might be "useful" to carry out studies involving certain variables. Both authors demonstrate a positive attitude towards the value of particular research.

4.1.2.3 Describing the Conceptual Model

Table 4.1 shows that one sixth of the introductions employed a *Describing the conceptual model (IDM)* move (17.5%; f = 52; n = 21). After a niche was established, the writers might devote several paragraphs to a description of the conceptual framework underlying the research question. In the descriptions, based on

a theoretical framework resultant from a thorough review of literature organized around the topic, a conceptual framework specified the relations among theoretical constructs or variables identified in the study. In addition, it outlined the input, process and output of the whole investigation.

In the ERC introduction sections, conceptual models often stood out from the rest of the text. They were presented under sub-headings containing terms such as "conceptual framework" and "conceptual model". Otherwise, they came under content-specific sub-headings, for instance, "Effective Reading Instruction."

> *26) <IDM1> This process evaluation adapted the conceptual framework described by Linnan and Steckler. </IDM1> <IDM2> Following an extensive review of process evaluation research, they developed a framework for systematically conducting process evaluations to capture key process components: context, reach, dose delivered, dose received and fidelity of intervention delivery.</IDM2> <IDM3> 'Context' assesses aspects of the intervention's environment or setting that might affect its delivery or outcomes. 'Reach' is the proportion of the intended audience who actually took part in the intervention. 'Dose delivered' is the proportion of the intended intervention that was actually provided to the target audience. 'Dose received' is the extent to which the intended audience engaged with and used the intervention information and activities. 'Fidelity' is the quality of the intervention implementation, the extent to which it was delivered in the intended manner. Linnan and Steckler additionally discussed assessing recruitment and combining reach, dose delivered, dose received and fidelity to create a composite implementation score. We could not analyze recruitment as a key component of this process evaluation because we did not collect data on reasons for nonparticipation. We also did not create a composite implementation score because of the conceptual difficulty of determining the relative weight of each component when combined.</IDM3> .... (HER_09)*

Often, as demonstrated by the unabbreviated excerpt, authors took three steps to present a conceptual framework. Step 1, *Introducing the conceptual framework (IDM1)*, named the framework. Step 2, *Identifying the components (IDM2)*,

listed its components. Details of each component and their relationships were provided in Step 3, *Describing components' relationships (IDM3)*.

These three steps were observed to occur in the order of IDM1, IDM2 and IDM3. IDM1 and IDM3 were nearly always used together while in over half of the cases, IDM2 were absent because the components were identified either in IDM1 or in IDM2.

4.1.2.4 Presenting the Present Research

The final move of the introduction section, *Presenting the present research (IPR)*, had the second largest number of observations (f = 497) and were present in all the RAs, hence an obligatory move (see Table 4.1). In Swales' terms (1990), IPR occupies the niche identified by IEN and turns it into a research space that the present study fits right in. The greatest majority of IPRs (90.8%) followed an *Establishing a niche (IEN)* move, either filling in a specific gap just established in the course of reviewing past research or addressing the gap at the close of the introduction. Nevertheless, seven fronted IPRs were observed in the introduction sections.

In the articles under investigation, IPRs were realized by one or a combination of the following steps: *Announcing the present research (IPR1)*, *Announcing principal findings (IPR2)*, *Indicating RA structure (IPR3)*, *Claiming significance of the research (IPR4)*, *Verifying compliance with ethical standards* (IPR5), and *Describing research context (IPR6)*.

4.1.2.4.1 Announcing the Present Research

As can be seen in Table 4.1, *Announcing the present research (IPR1)* far outweighed the other possible steps both in frequency and distribution (99.2%; f = 392; n = 119), qualifying itself as the only conventional step of the *Presenting the present research (IPR)* move.

Two types of IPR1s were observed in the corpus. General IPR1s, always placed towards the end of an introduction, announced the present study in summary form, simultaneously dealing with several aspects of the study such as purpose, research design, hypotheses, and research questions. Without exception, all the RAs containing an IPR1 step contained a general IPR1. The following excerpts illustrate this type of IPR1s.

27) In an effort to determine ..., the present article uses ... data to examine the effects of ... To extend the literature, we address ... in two new ways. First, we minimize ... by ... Second, we consider whether ... .(Abbreviated) (SOE_06)

28) Thus, the purpose of our study was to examine the association of home and parental factors and child's preferences with AA of FV in the home environment of elementary-aged Hispanic children. In this study, home and parental factors included parental practices that promote FV intake, parental role modeling, perceived cost of FV, perceived benefits of fast food and amount of TV viewing. .... (HER_01)

By contrast, many writers used specific IPR1s to introduce one particular aspect of the current study with regards to a specific research gap or deficiency exposed in the IEN move. From Table 4.1, we can infer that approximately 70% of the IPR1s identified belong in this category. In fact, at least half of the RAs in

the corpus were found to contain such IPR1s scattered across the introduction section.

29) This study will help to address this hole in the literature. Using the NICHD SECCYD dataset, we were able to perform one of the first comparisons of classroom processes in public and private schools. ....(ESJ_4)

30)Thus, we address the following research question: RQ3: To what extent will a video's message characteristics (MSV and message appeals) be associated with viewer responses (i.e., number of viewers, number of comments, and viewer rating)?... (JEE_10)

As revealed by the data, specific IPR1s could be about gaps in different aspects of the study ranging from the research deign, to research questions, to research purposes, to hypotheses. Excerpt 29 addresses a "hole," just identified in previous research, which is the absence of comparative studies of classroom processes. In a similar vein, Excerpt 30 raises a specific research question that has not been addressed previously.

4.1.2.4.2 Announcing Principal Findings

Table 4.1 indicates that *Announcing principal findings* (IPR2) was an optional step (6.7%; f = 8; n = 8) which was thinly distributed in the corpus. By this step, a few writers highlighted their major findings to arouse interest in the current study, as illustrated in the following excerpts.

31) Our results indicate that .... We also find that race and the percentage of minority students within a school are related to a student's likelihood of extracurricular participation. Overall, participation in EAs explains a modest portion of the SES advantage in both noncognitive and cognitive skills. Finally, the association between extracurricular participation on noncognitive and cognitive skills depends in part on students' SES. ....(Abbreviated)(SOE_02)

32) The results of this study indicate that many interdisciplinary graduate program proposals lack strong connections between desired outcomes,

evidence, and learning experiences, and suggest that more thorough integration of
each of these goals may better support new and continuing interdisciplinary programs.
... (JEE_08)

4.1.2.4.3 Indicating RA Structure

As shown in Table 4.1, *Indicating RA structure (IPR3)* was an

optional step (30.8%; f = 43; n = 37) in the ERC corpus.

33) In what follows, an experimental study is presented in which
the effects of text segmentation, picture labeling, and the combination of both were
investigated. Furthermore, learning from spatially separated texts and pictures was
compared to learning from a spatially integrated text and picture. A discussion of the
theoretical as well as the practical implications of the findings of the study is given at
the end. .... (LAI_04)

34) The remainder of this article is structured as follows. We begin
with .., followed by ... To delineate ... we attempt to identify ... Then we build ... and
develop a preliminary framework ... This is followed by an empirical analysis of two
cases ... The article concludes with a discussion of ... . (Abbreviated) (JLS_03)

As the above excerpts show, this step, invariably placed at the

close of an introduction, informed the reader what came after and/or how the content

of the rest of the article was organized.

4.1.2.4.4 Claiming Significance of the Research

Claiming significance of the research (IPR4) was found an optional

step (22.5%; f = 34; n = 27) in presenting the present research. The writers might choose

to claim assertively the value of their research hoping to impress the reader and attract

him or her to the details of the study presented in the ensuing sections.

35) By exploring this question, our contribution is threefold. First,
we introduce…, entailing important implications for ... Second, we address the
assertion ... To address this issue, we use a sample of ... The results of the analysis
provide support for …, leading to important implications for ... Finally, our study

moves beyond ... By doing so, it empirically examines…(Abbreviated) (LAE_09)

36) Our findings have implications for the theory of education for sustainability as well as the practice of implementing sustainability within the business school, the university, and the wider community. (LAE_08)

The excerpts above indicate that the writers might highlight theoretical or practical values of their research findings. Other aspects highlighted ranged from the ingenuity of the research design, to the uniqueness of the research topic, to the inspiration the research might give to other researchers.

4.1.2.4.5 Verifying Compliance with Ethical Standards

One introduction section was observed to contain a *Verifying compliance with ethical standards (IPR5)* step, which has seldom been reported in previous literature.

37) The naturally occurring reputation manipulation avoided the unethical proposition of purposefully inducing a negative bias into an actual work setting...(LAE_06)

The excerpt above was taken from an article reporting a field experiment that investigated how a negative reputation of teachers biased student evaluations of teaching. The author clears possible ethical concerns by declaring that the research was designed and conducted in no way that would actually damage any participant's reputation.

4.1.2.4.6 Describing Research Context

Table 4.1 shows another optional step, *Describing research context (IPR6)*, of the *Presenting the present research (IPR)* move. This step, found in

a small number of articles in the corpus (14.2%; f = 19; n = 17), occurred mostly

immediately after an *Announcing the present research (IPR1)* step. The background

information provided was directly related to the study with reference to the research

site, participants, curriculum, intervention program, etc.

> *38) In this section, we provide some related contextual information*
> *about our corpus. One of these factors is ...'*

> *In both first-year classes in our corpus, the teacher stood at the*
> *front with the students in rows...*

> *Another powerful set of factors to consider involves differences in...*

> *Another primary factor relates to the knowledge base of...*

> *Fourth-year students may well know…, but this does not mean that*
> *hat first-year students know nothing....................................(Abbreviated)   (CAE_07)*

> *39) ANTECATALYST was used in the curriculum of second-year*
> *psychology students in 2008. ANTECATALYST took the students through a 3D North*
> *African saltpan to learn about the navigational strategies of Cataglyphis ants. Three*
> *groups learned with flat, cartoon, or lifelike AVAs. ....(CAE_07)*

Excerpt 38 was taken from an article that investigates classroom

talk using a digital corpus of recordings of psychology classes. Immediately after

announcing the present research (IPR1) and before the methods section, the author

describes in five paragraphs the context in which the classes were taught. The

information covered is comprehensive — seating arrangements, differences in

pedagogical goals, scope of topics, course readings, student evaluation, and students'

knowledge base. Incidentally, this detailed description was found under the

sub-heading "Contextual Considerations."

Similarly, in Excerpt 39, the author, who investigated the impacts of animated-virtual actors' visual complexity, provides information about the curriculum involved in the investigation with regards to materials used and the way different participating groups were taught. This particular chunk of text appeared early in the article as the author announced the present research before establishing a niche.

4.1.2.5 Move Cyclicity and Generic Rhetorical Structure

Move recycling was found to be a remarkable characteristic of the introduction sections investigated. Table 4.1 suggests that the most cyclic move was *Establishing a territory (IET)*, which recurred 11 times on average in the articles. *Presenting the present research (IPR)* was observed four times per text on average. Less cyclic moves were *Establishing a niche (IEN)* and *Describing the conceptual model* (IDM), whose average recurrence rates were three and two, respectively.

As indicated in 3.3.4, to explore what patterns moves most likely form with other moves, salient multi-move sequences with both a minimum frequency and a minimum range of 35 were extracted using the n-gram function of AntConc (Anthony, 2014). The same operation was repeated in analyzing the move structures of the other three RA sections. Table 4.2 below lists multi-move sequences thus obtained in the introductions. P-values indicate the level of confidence with which we judge how strongly the first move is associated with the rest of the sequence.

Clearly, distinct move sequences involved the three obligatory or conventional moves while the optional move *Describing the conceptual model* did not seem to pattern with other moves due to its low frequency of occurrence. The most notable were the 2-move patterns IET-IEN, IPR-IET, IET-IPR and IEN-IPR, which occurred conspicuously frequently in large numbers of the articles. Patterns consisting of 3, 4, and 5 moves also abounded. Among them was the classic 3-move CARS model IET-IEN-IPR, which was observed with a frequency as high as 130 in over two thirds of the total number of articles. The 4-move pattern IET-IEN-IET-IEN was identified in over half of the corpus. The still more complex 5-move pattern IET-IEN-IET-IEN-IPR appeared in over one thirds of the corpus, and it recurred more than once in a number of articles.

Table 4.2 Multi-move Sequences in the Introduction Section

Multi-move sequence	Frequency	Range	Probability	Rank
IET IEN	299	112	0.10	1
IPR IET	218	96	0.12	2
IET IPR	199	86	0.06	3
IEN IET	183	89	0.09	4
IET IEN IET	162	86	0.05	5
IEN IPR	154	90	0.08	6
IET IEN IPR	130	84	0.04	7
IET IPR IET	123	60	0.04	8
IEN IET IEN	117	67	0.06	9
IPR IET IPR	105	50	0.06	10
IET IEN IET IEN	104	64	0.03	11
IPR IET IEN	102	70	0.06	12
IEN IPR IET	88	68	0.04	13
IET IEN IPR IET	69	55	0.02	14
IET IPR IET IPR	69	37	0.02	15
IPR IET IPR IET	67	36	0.04	16
IPR IET IEN IET	62	51	0.03	17
IEN IET IPR	60	48	0.03	18
IEN IET IEN IET	57	40	0.03	19
IEN IET IEN IPR	56	47	0.03	20
IET IEN IET IPR	53	44	0.02	21
IET IEN IET IEN IET	52	37	0.02	22
IEN IPR IET IEN	50	41	0.03	23
IET IPR IET IEN	49	41	0.02	24
IET IEN IET IEN IPR	48	44	0.02	25
IPR IET IEN IET IEN	38	36	0.02	26

The high frequency patterns observed also suggest that move cycling did not just involve individual moves. Sequences of moves in the introduction texts, particularly 2-move sequences, recycled as well. More interestingly, variants of the basic IET-IEN-IPR pattern can be discerned from quite a few introduction texts.

Regarding the order of appearance of the moves, it was found that an overwhelming number of introductions opened with IET (109 out of 120). The remaining few introduction texts had a fronted IEN or IPR. It was also observed that all texts ended with an IPR.

From the results presented above, a generic move structure of the introduction section emerges. Thus, the scenario of the this section unfolds like this: Writers of education research open an introduction by situating their own research in a relatively broad area of research through claiming how important the topic is (IET2) and what is currently known (IET3), or reviewing relevant past research (IET4), or defining terms and concepts (IET5). Next, they proceed to establish a research space (IEN). To do so, they need to pinpoint a gap in previous research (IEN1B). They may also claim that previous research (IEN1A) is problematic in some way, or raise questions about unresolved issues (IEN1C), or indicate a need to expand knowledge on the topic (IEN1D). Alternatively, a research space may be established with positive reasons for a particular study (IEN2). They may go through the *Establishing a territory‑Establishing a niche* cycle a few times, moving from the most general to the least

general. Subsequently, writers may describe the conceptual framework within which they execute their study (IDM). If they choose to do so, they need to name the framework (IDM1), indentify the elements (IDM2) and explain how the elements are mutually related (IDM3). Then comes the most exciting moment of previewing their own research that fits right in the space they have just created (IPR). They announce their research with respect to purposes, design, research questions, or hypotheses (IPR1). They may also occasionally preview their major findings. To provide transition, writers may end the section showing the reader how the rest of the article proceeds (IPR3). However, prior to that, they may highlight the significance of their research (IPR4) or provide contextual information about the research site, participants, curriculum, intervention program, etc. (IPR6).

4.1.3 The Methods Section

Table 4.3 Rhetorical Move Structure of the Methods Section

Code	Move	f	n	%	S
MRD	Describing research design	69	62	52	Opt
MDC	Describing data collection procedures	625	12	100	Obl
MDC1	Describing research context	75	44	36.7	Opt
MDC2	Describing the sample	140	79	65.8	Con
MDC3	Describing instruments	199	56	46.7	Opt
MDC4	Elaborating on data collection procedures	162	57	47.5	Opt
MDC5	Justifying data collection procedures	30	9	7.5	Opt
MDC6	Verifying compliance with ethical standards	19	17	14.2	Opt
MDA	Describing data analysis procedures	176	93	77.5	Con
MDA1	Recounting data analysis procedures	125	64	53.3	Opt
MDA2	Justifying data analysis procedures	17	14	11.7	Opt
MDA3	Establishing inter-coder reliability	34	26	21.7	Opt

Note. f = frequency of a move/step in all the RAs; n = number of RAs in which a move/step appears; % = percentage of a move/step in all the RAs; S = Status; Obl = obligatory; Con = conventional; Opt = optional

4.1.3.1 Describing Research Design

As seen in Table 4.3, *Describing research design (MRD)* was an optional move in the methods sections (52%; f = 69; n = 62). This move clarified some vital aspects of the research design ﹘ research objectives, type of research design, hypotheses, or variables involved. As expected, this move was found to always precede the other two moves in the methods sections.

> *40) This study employed a quasi-experimental design for cross-sectional data in which the high school mathematics curriculum (Group) a student completed was the independent variable of most interest. Difficulty of students' first university-level mathematics course and the grade earned in that course served as dependent variables. ....(AER_02)*

> *41) The goal of this experiment was to assess the effect of... The reason for sparing the high relevance condition was that ... The Experiment 1 used a 2 x 2 between subjects design, whereby the effects of the two experimental factors quality information and content relevance were observed with respect to students' final achievement outcome. It was expected (Hypothesis 1) that ... (Abbreviated) (LAI_09)*

Excerpt 40 is a succinct description of a research design, with its type identified as quasi-experimental; high school mathematics curriculum as the independent variable; and difficulty of the first college mathematics course and grade earned as dependent variables. Excerpt 41 is an abbreviated excerpt from an article reporting a study involving two experiments. It covers several aspects of the first experiment: its objective (to assess the effect), its type (a 2 x 2 experiment), rationale for excluding a possible independent variable (the reason for sparing), and expected results (hypotheses).

4.1.3.2 Describing Data Collection Procedures

As Table 4.3 shows, *Describing data collection procedures (MDC)* was present in all methods texts (100%; f =625; n = 120). In cases where *Describing research design (MRD)* was absent, this move was always the first in the methods sections. The data collection move was usually realized by a combination of some of these constituent steps: *Describing research context (MDC1)*, *Describing the sample (MDC2)*, *Describing instruments (MDC3)*, *Elaborating on data collection procedures (MDC4)*, *Justifying data collection procedures (MDC5)*, and *Verifying compliance with ethical standards (MDC6)*.

4.1.3.2.1 Describing Research Context

A *Describing research context (MDC1)* step was found optional (36.7%; f = 75; n = 44) in the methods sections. Identical with the *Describing research context (IPR6)* step of the *Presenting the present research (IPR)* move in the introduction sections, MDC1 performed the function of providing relevant background information for the research. MDC1 texts varied in length, ranging from one single sentence to extensive paragraphs depending on the number of topics covered.

> *42) I selected Midwestern University (MU), a highly regarded public research university ..., as the site for this study. MU houses both an undergraduate elementary education program that ... and an undergraduate secondary social studies education program ... In each program, prospective teachers become certified to teach social studies in the middle grades. ....(Abbreviated) (AER_05)*

43) The Crystal Springs School District is a large elementary school district in Southern California. In 2003-2004, the district had 40 elementary schools and over 1000 teachers, with over 90% holding full teaching credentials. The district serves a diverse student population including more than 25, 000 students. In 2003-2004, the student population was 64.3% Latino or Hispanic, 17.0% White (not Hispanic), 8.5% Filipino, 4.9% African American, 3.6% Asian, 0.9% Pacific Islander, and 0.4% American Indian or Alaskan Native. Additionally, about half of the students in the district were eligible for free or reduced-price lunch, and more than one third of the students were English language learners.. ....(AER_07)

MDC1 usually covered a wide range of information categories. Excerpt 42 describes the research site and existing educational programs. In considerable detail, Excerpt 43 provides a broader context in which the study took place, including the geographical location of the research site, the school system, teacher certification, an ethnically diversified student population with a low socio-economic status. Other types of information was also noted in the corpus, for example, educational policies, on-going intervention programs, and public attitudes.

4.1.3.2.2 Describing the Sample

As Table 4.3 shows, *Describing the sample (MDC2)* occurred quite frequently   (65.8%; f = 140; n = 79), qualifying itself as an conventional step. Unlike *Describing research context (MDC1)*, which described the general context of a study including characteristics of the general population of a community, MDC2 aimed to provide an adequate description of the sampling procedure and the major characteristics of the sample often called participants in education research.

44) Participants in the study included children and their families. In 1991 NICHD researchers recruited mothers with newborns from hospitals located at 10 different geographic sites across the United States. Out of the 8, 986 mothers

whom the researchers initially contacted, 5,416 met eligibility criteria and agreed to be contacted after they had left the hospital. Of these, a randomly selected subgroup was chosen with procedures to ensure ethnic, economic, and educational diversity, resulting in 1,364 families with healthy newborns originally enrolled in the study. Although there was attrition over the years, researchers followed the remaining participants through adolescence. ....(ESJ_04)

45)   Forty-one kindergarten children, recruited from an urban elementary school located in a southeastern state, served as participants. Of the 24 female and 17 male participants, 36 were Caucasian (88%), 3 were African American (7%), and 2 (5%) were either multiracial or listed as "other ethnicity ." Four of the 41 children reportedly spoke a second language, although English was their first language. Information regarding a second language was not available for 3 other children. The children ranged in age from 5.8 to 6.7 (M = 6.2).

According to school records, 7 % of the children in the school qualified for free and reduced-price lunch status, suggesting students were primarily from middle-class to upper middle-class homes.

According to teacher report, all participants had negative histories for speech, language, motoric, sensory, or intellectual deficits.

In addition, all children scored above a standard score of 80 on the Kaufman Brief Intelligence Test-2 confirming nonverbal cognitive skills within typical limits. ....(SSR_08)

As a rule, as the above excerpts demonstrate, MDC1s in the methods texts described the sample and the sampling method as specifically as possible, focusing on characteristics that might have some bearing on the interpretation of results. Often, besides sample size, information might also be provided about age, gender, ethnic and/or racial group, level of academic attainment, socioeconomic status, disability status, etc.

4.1.3.2.3 Describing Instruments

Another step found common in the corpus was *Describing instruments (MDC3)* (46.7%; f = 199; n = 56). By this optional step, some writers

described in great detail the instruments, e.g., questionnaires, interviews, or tests, for

collecting data needed to measure the variables included in the research design.

46) The focus interviews focused on a number of themes, namely: (1) what position ICT and media had in the teachers' subjects, (2) how they thought that the use of media and ICT influenced the working method in the classroom, (3) the relations to the pupils, (4) the subject content itself and finally (5) the role of the teacher. ....(CAE_04)

47) To assess reading ability, the Word Attack and Word Identification subtests of the WRMT-R were administered. Raw scores on the two subtests were summed to create a reading composite score. Spelling ability was assessed using Wolter and Apel's (2010) 10-item, spelling dictation task: "fan," "pet," "dig," "mop," "rope," "wait," "chunk," "sled," "stick," "shine." The children's responses were scored using Wolter and Apel's 8-point rating scale. Interscorer agreement was calculated for the four word learning tasks and the spelling task for 9 (22%) participants. Agreement was calculated at 100% for all word learning tasks and 95% for the spelling dictation task. ....(SSR_08)

While Excerpt 46 identifies the data collection instrument as focus

interview and lists questions to guide the interviews, Excerpt 47 explains variables,

the use of available instruments, their validation and the scoring method.

4.1.3.2.4 Elaborating on Data Collection Procedures

Elaborating on data collection procedures (MDC4) was another

optional step almost as important as MDC3 (47.5%; f = 162; n= 57). MDC4s gave a

step-by-step description of the delivery of instruments to the participants. As to their

position, they appeared either before or after an instrument description step.

48) A post-intervention quiz was administered to both the Control and Experimental groups at the end of the intervention period. .... (JEE_05)

49) All participants were administered the experimental spoken and written word learning procedure first. Following this administration, other language and literacy tasks were conducted in randomized order. Children were

tested across 2 days, typically within 1 week's time. Each session lasted approximately 45 min. All tasks were administered by undergraduate or graduate students trained in all procedures used in the study by the author. ....(SSR_08)

Excerpt 48 states what instrument was applied, to whom it was applied, and when it was applied. As the indefinite article preceding "post-intervention quiz" indicates, this step is followed by an MDC3 step that describes the quiz as an instrument.

Excerpt 49 provides a chronological narrative of the procedures of applying the instruments to the participants with details on what instruments were applied, who administered them, how the participants were grouped, and how much time was allocated. The definite article preceding "experimental...procedures" indicates the step's position after an instrument description step.

4.1.3.2.5 Justifying Data Collection Procedures

In comparison with the other steps, *Justifying data collection procedures (MDC5)* was observed significantly less frequently in the corpus (7.5%; f = 30; n = 9). A few writers took this step to establish the appropriateness of the data collection procedures. The purpose was to defend the study against any doubt regarding validity and reliability.

50) We did not include these features in the instruction because we wanted to see if the students would extend what they learned about words to novel letter strings (i.e., the nonwords). .... (SSR_07)

51) We opted for a group size of nine, since research of Schellens and Valcke (2006, p. 349) showed that groups of this size perform better than larger groups and that "discussion in small groups reflects larger proportions of higher levels of knowledge construction".. (LAI_05)

As the above excerpts show, important decisions concerning data collection were necessarily rationalized. Excerpt 50 explains why in the course of the intervention the researchers never told the participants that the pronunciation of known words could help pronounce new words with similar spellings and why they never used the term "strategy." Likewise, Excerpt 51 cites previous research to support the researcher's decision on group size when group discussions were held for data collection. Other instances of this step in the corpus concerned sample size, instruments, timing of activities, and other aspects of data collection.

4.1.3.2.6 Verifying Compliance with Ethical Standards

Interestingly, while a single case of *Verifying compliance with ethical standards (IPR5)* step was found of the *Presenting the present research (IPR)* move in the introduction sections, the same strategy was used significantly more frequently as a step realizing the *Describing data collection procedures (MDC)* move in the methods sections (14.2%; f = 19; n = 17).

> 52) *The University of Western Australia Human Research Ethics Committee approved the study protocols.* .... *(HER_02)*

> 53) *Since this research took place in a naturalistic setting, depriving students completely of tutor support was considered unethical, so a condition in which students never received peer tutor support was not included.* .... *(LAI_05)*

4.1.3.3 Describing Data Analysis Procedures

Table 4.3 shows that *Describing data analysis procedures (MDA)* was an conventional move in the methods sections of the ERC corpus (77.5%; f = 176; n =

93). Almost always following a *Describing data collection (MDC)* move, this move described in varying degrees of detail how data were analyzed for results in answer to the research questions, which is one of the fundamental aspects of the conduct of research. This communicative purpose was realized by one or a combination of three steps, viz., *Recounting data analysis procedures (MDA1)*, *Justifying data analysis procedures (MDA2)* and *Establishing inter-coder reliability (MDA3)*. Details of the steps are presented hereafter.

4.1.3.3.1 Recounting Data Analysis Procedures

Recounting data analysis procedures (MDA1) was found to be the most frequent step of the *Describing data analysis (MDA)* move (53.3%; f = 125; n = 64). Over half of the writers took this optional step to narrate the particulars of the analytical treatment of data, be it quantitative or qualitative.

> 54) *In order to examine whether or not public and private schools differ in terms of classroom characteristics and processes we used one-way ANOVA tests. We did not apply the Bonferroni correction, purposefully opening up the possibility of introducing Type I error in an effort to avoid prematurely embracing the null hypothesis of no differences. ....(ESJ_04)*

> 55) *Our data analysis started with the semi-structured interviews... We followed an analysis approach similar to the methods described by Miles and Huberman (Citation). First, the interviews were grouped by case… Then, each case was analyzed separately before looking across cases for themes and patterns. Our analysis included a combination of coding output from Atlas Ti and pictorial and graphical data representations (Citation). Miles and Huberman (Citation) describe three categories of data analysis strategies including: a) case-oriented… b) variable-oriented... and c) mixed strategies... This study incorporated a mixed strategy where the eleven individual cases and the four STV categories (value constructs) were equally important. Results are presented as variable-oriented assertions such that they are organized around the four STV categories including attainment, cost, interest, and utility. .... (Abbreviated) (JEE_06)*

Excerpt 54, from a quantitative study, explains what analytic method was used to answer what specific research question. In addition, it gives a reason why another procedure was not applied.

Excerpt 55 was taken from a qualitative study. It reviews how interview transcripts were thematically coded. Details include what coding scheme was used, how data was grouped, what software programs were used, what analytic strategy was employed, and how results were presented.

A comparison of the two excerpts reveals a marked difference in the use of citations. Excerpt 54 is without citations, indicating the author's assumption that the statistical methods mentioned are part of the basic professional knowledge of the discourse community. In contrast, Excerpt 55 is heavily loaded with citations because the author does not presume that the readership have prior knowledge of the coding scheme, the software programs and the analytic strategy mentioned. Indeed, the excerpts represent two types of *Recounting data analysis procedures (MDA1)* steps in the corpus.

4.1.3.3.2 Justifying Data Analysis Procedures

Another optional step, *Justifying data analysis procedures (MDA2)*, was observed in the methods sections (11.7%; f = 17; n = 14). In the texts investigated, this step established the appropriateness and vigor of the procedures used in approaching the data for results. Like the *Justifying data collection procedures*

(MDC5) step, it contributed to the validity and reliability of research results. As anticipated, this step usually followed *Recounting data analysis procedures (MDA1).* The following excepts exemplify this step.

> *56) Due to the large sample size, nearly all differences, even minute and meaningless differences, would have statistical significance at p < 0.05. Therefore, we used the eta-squared test of significance (Citation). Eta-Square is a measure of strength of association between variables. It represents the proportion of dependent variable variance accounted for by any factor of interest. It is independent of sample size and ranges from 0 to 1 (Citation). The eta-square is conventionally defined as a small effect is 0.01, a medium effect is 0. 06, and a large effect is 0.14 (Citation).* ....*(JEE_07)*

> *57) It has been suggested that inclusion of the naming duration could be a more sensitive and reliable measure of children's decoding efficiency (Citation). Due to the reading instruction methods in Finland, children learn to vocalize their decoding substages. Some poor readers tend to blend phonemes in a serial manner, and the onset time does not capture this process.* ....*(SSR_05)*

As can be seen, both excerpts give persuasive reasons for important decisions made in the data analysis phase. Excerpt 56 not only explains why a large sample necessitates the use of eta-squared test of significance but also introduces this analysis in terms of effectiveness. Likewise, Excerpt 57 provides a reason for the inclusion of the naming duration as a variable in the analysis.

Inevitably, MDA2 texts were heavily loaded with citations. Without authoritative references cited, the reader might not be convinced of the appropriateness and robustness of the analytic method used.

4.1.3.3.3 Establishing Inter-coder Reliability

Different from previous research findings, the corpus revealed an optional

Establishing inter-coder reliability (MDA3) step (21.7%; f = 34; n = 26). Taking this

step, the authors demonstrated that their analyses were satisfactorily accurate and

reliable, hence the results were trustworthy.

> *58) In the course of developing our coding scheme, we continuously monitored inter-coder reliability. Reliability was not a major problem, since our coding scheme was straightforward and factual in character and called for little difficult interpretation..(SOE_04)*

> *59) Two coders independently rated 10% of the data for each coding schema. Reliability was established for the Triadic/Non-triadic schema with a Cohen's kappa of 0.87 and for the Teacher Questions schema with a kappa of 0.90. Remaining disagreements were resolved by discussion................................. (RST_03)*

Both authors of the above excerpts try to convince the reader that

their analyses were not problematic as multiple coders analyzed the same data and a

high level of agreement was reached.

4.1.3.4 Move Cyclicity and Generic Rhetorical Structure

The methods sections were found to be the least cyclic among the four

RA sections in the ERC corpus. As Table 4.3 shows, in the methods sections, two

moves tended to recur. The average recurrence rates per article of the *Describing data*

collection procedures (MDC) and the *Describing data analysis procedures (MDA)*

moves were five and two, respectively. The other move, *Describing research design*

(MRD), hardly recycled. Yet, unlike in the other sections, multi-move sequences

recycled just occasionally in a limited number of articles.

Concerning the move structure of the methods sections, three important

patterns are displayed in Table 4.4. The MDC-MDA pattern was the most prominent,

appearing at least once in 89 out of 120 articles. The MRD-MDC pattern was observed in 52 articles. Even the more involved pattern MRD-MDC-MDA appeared in one third of the articles.

Table 4.4 Multi-move Sequences in the Methods Section

Multi-move sequence	Frequency	Range	Probability	Rank
MDC MDA	99	89	0.45	1
MRD MDC	60	52	0.42	2
MRD MDC MDA	40	39	0.28	3

The results also indicate that in the methods sections, if *Describing research design (MRD)* was present, it was always the opening move. If not, all but one methods section began with *Describing data collection procedure* (MDC) and ended with *Describing data analysis procedures (MDA)*.

From these results, an optimal generic move structure of the methods section was derived (see Table 4.3). Accordingly, writers may begin a methods section by describing the research design (MRD). Then, they describe procedures of data collection (MDC), providing details of the research context (MDC1), the sample (MDC2), instruments (MDC3), actual procedures (MDC4), justifications for the procedures (MDC5) or proof of compliance with ethical standards (MDC6). Afterwards, they need to describe procedures of data analysis (MDA), explaining how they analyzed data (MDA1) and why they analyzed the data the way they did (MDA2). Occasionally, they may need to indicate inter-coder reliability if the data were analyzed manually (MDA3).

4.1.4 The Results Section

Table 4.5 displays descriptive statistics of the moves and steps observed in the results sections. As can be seen, 4 moves were observed in the results sections. They are *Providing background information (RBI)*, *Reporting specific results (RRR)*, *Commenting on specific results (RCR)*, and *Summarizing results (RSR)*. RRR was the only obligatory move. Based on the cut-off point of 60%, RCR was identified as conventional while RBI and RSR optional.

Table 4.5 Rhetorical Move Structure of the Results Section

Code	Move	f	n	%	S
RBI	Providing background information	88	67	56	Opt
RRR	Reporting specific results	2158	120	100	Obl
RRR1	Preparing for specific results presentation	562	106	88.3	Con
RRR2	Presenting results	1180	120	100	Obl
RRR3	Illustrating with examples or excerpts	416	55	45.8	Opt
RCR	Commenting on specific results	274	84	70	Con
RSR	Summarizing results	62	29	24.2	Opt

Note. f = frequency of a move/step in all the RAs; n = number of RAs in which a move/step appears; % = percentage of a move/step in all the RAs; S = Status; Obl = obligatory; Con = conventional; Opt = optional

Findings about each move and step are reported separately in detail hereafter.

4.1.4.1 Providing Background Information

As Table 4.5 shows, *Providing background information* (RBI) was an optional move bordering on the conventional status (56%; f = 88; n = 67). As such, RBI, unless not present, always initiated results reporting and prepared the reader for

the long-awaited results of the research. This communicative purpose was achieved

mostly by reiterating the general research purpose and the data analysis procedures or

occasionally by providing descriptive statistics of the sample.

60) For each of our two protocols, we first tested the normality of the distribution and then used exploratory factor analysis to propose an underlying factor structure. We used confirmatory factor analysis to test the relationships between our observed variables and underlying latent constructs. Finally, we used structural equation modeling to explore relationships between coaching factors and instruction factors. ....(ESJ_10)

61) Descriptive statistics for all variables are presented in Table 2. Correlations were low to moderate. This suggests our measures assessed different constructs and warranted regression analyzes. In an initial data screening, we found no violations of regression assumptions or multicollinearity. As mentioned above, all correlations with age-related epistemic beliefs (EB Age) were calculated only for the age group of 51 to 65 years; the respective descriptive statistics are presented in Table 3.. (LAI_06)

Excerpt 60 restates briefly what statistical analytic methods were used for

data analysis. This reminder is intended to help the reader better comprehend and

appreciate the results soon to be reported. For the same purpose, Excerpt 61 provides

descriptive statistics and preliminary results based on them.

Otherwise, the communicative purpose of this move might be achieved

by establishing trustworthiness of data collection, outlining results, and providing

textual transition.

4.1.4.2 Reporting Specific Results

Quite expectedly, *Reporting specific results* (RRR) was by far the most

dominant move in the results sections of the ERC articles (100%; f = 2158; n = 120).

In this central, obligatory move, the writers presented their results or findings with accuracy and clarity in correspondence with the research questions or different instruments that generated the results. As Table 4.5 demonstrates, this communicative purpose was fulfilled by a certain configuration of these three steps: *Preparing for specific result presentation (RRR1)*, *Presenting results (RRR2)* and *Illustrating with examples or excerpts (RRR3)*. Details of the steps are reported subsequently.

4.1.4.2.1 Preparing for Specific Results Presentation

Table 4.5 shows that *Preparing for specific results presentation (RRR1)* was highly conventional in the corpus (88.3%; f = 562; n = 106). Different from the *Preparing for results preparation (RBI)* move aimed at preparing the reader for the entire results section, this step was aimed at getting the reader ready for specific results. Logically, it always preceded the *Presenting results (RRR2)* step.

> *62) Our first research question concerned how often top-tier journal articles include explicit implications for practice...............................(LAE_04)*

> *63) Because some children never made miscues in all five categories, and some parents responded multiple times with the same feedback category to one particular type of child miscue, several paired comparisons were computed to determine if the type of parental feedback provided differed depending on the proceeding miscue...(SSR_06)*

Whereas Excerpt 62 reminds the reader of an individual research question immediately before a specific result is reported, Excerpt 63 explains what statistical analysis was performed and why it was used to answer an individual research question.

Alternatively, some writers might take this step to brief the reader on other things necessary for understanding a specific result, such as data treatment and participants' background.

4.1.4.2.2 Presenting Results

As expected, the *Presenting results* (RRR2) step had the highest frequency and the widest distribution among all moves and steps (100%; f = 1180; n = 120). Mostly following an RRR1, this step presented specific results in an objective and impersonal manner.

> *64) On the Spelling experimental measure in English, after adjusting for initial Letter-Word Identification performance, intervention students outperformed comparison students, F(1,104) = 5.41, p < .022, with a modest effect size (d = 0.40). ....(AER_08)*
>
> *65) In the teachers' logs two forms of frictions were reported: (a) a completely unexpected event (positive or negative) took place, and (b) the realization that a teacher's usual teaching approach did not work any longer. The former always took the form of an incident, the latter usually referred to a longer period of time during which there was a growing feeling of discontent. Experiencing friction usually was an incentive for other learning activities such as considering own practice (see above) and/or experimenting. However, experiencing friction was also reported as a distinct learning activity that led to awareness of a discrepancy and/or a strong emotion, such as disappointment or pride, related to the discrepancy. Teachers experienced friction individually or with others. ....(LAI_10)*

As demonstrated in Excerpt 64, in quantitative studies results were clearly stated along with specific statistics such as degree of freedom, significance level, and effect size. In qualitative inquires, as shown in Excerpt 65, results were often in the form of themes and categories that had emerged from the data. Though not shown in the excerpt, quite often descriptive statistics such as frequencies, means and percentages were given for qualitative results.

4.1.4.2.3 Illustrating with Examples or Excerpts

This step was found optional in the results sections of the ERC articles (45.8%; f = 416; n = 55). Its communicative function was to ensure clear understanding of a specific result or to provide evidence for it.

66) A first-grade teacher described a student with a clinically diagnosed food obsession: "<EX>." Other examples of deviant student behaviors included stealing and compulsive lying. ....(ESJ_07)

67) For example, AIOs generally reported that they already had substantial knowledge of all their schools, including their new small autonomous schools, and that while they received various unsolicited information from the SSO administrators, that information did not build their knowledge of participating schools.. (AER_04)

In Excerpt 66 the author first quotes a particular teacher describing a student with food obsession as an example of deviant behavior, and then lists more types of such behavior. Obviously, the author's purpose is to clarify a finding already stated about student deviant behavior. However, Excerpt 67 is in itself a piece of evidence to support a finding or claim about AIOs' prior knowledge of their schools.

4.1.4.3 Commenting on Specific Results

Table 4.5 reveals that *Commenting on specific results (RCR)* was an conventional move (70%; f = 274; n = 84). Writers were found to compare their results with those of previous research or with their own anticipated results, or to interpret them, or to explain them, particularly if unexpected; or in some rare cases, to raise further questions. The main purpose was to offer their own opinions about the objective results.

68) If anything, these few significant results suggest that classroom processes may be of higher quality at public schools, at least in terms of instructional quality and amount of language arts instruction. .... *(ESJ_04)*

69) One possible reason why the estimated teacher effects on social and behavioral skill development are larger than the estimated teacher effects on academic development is that social and behavioral skills may not be measured as reliably as academic skills. .... *(SOE_05)*

70) This is in marked contrast to the earlier research with coaches (see, e.g., De Haan, 2008b, 2008c), where the participants nearly always described their own actions and their clients' responses before, during, and after the critical moment. .... *(LAE_10)*

While in Excerpt 68 the author makes sense of the findings, in Excerpt 69 the author accounts for the result just presented. Similarly, in Excerpt 70 the author states that the present result contradicts with those of other researchers. The following excerpt raises a question yet to be addressed.

71) What is unclear, however, is whether these secondary preservice teachers' learning could have been shaped differently if their methods course had explicitly emphasized challenging learners of all ages, including middle school students. .... *(AER_05)*

Interestingly, in commenting on their findings, the authors often did not hesitate to express their feelings, as demonstrated by the excerpt below.

72) This focus on teamwork is not surprising, given the emphasis on "team science" reflected in science and engineering practice, the literature (Citation), and engineering accreditation criteria. .... *(JEE_08)*

4.1.4.4 Summarizing Results

An optional *Summarizing results (RSR)* move was found in nearly a quarter of the results sections (24%; f = 62; n = 29). As the following excerpts demonstrate, RSR often concluded the section by recapitulating major findings and the authors' interpretations.

73) Together, these analyses suggest that the observed pattern of performance on phoneme awareness primarily reflected the influences of oral, phonological factors. ....*(SSR_09)*

74) In sum, the results indicate that the dyads in the collaborative video condition displayed a tendency to make fewer but more precise video selections and a tendency to change the order of the video segments more often than the dyads in the video player & text condition. .... *(JLS_07)*

4.1.4.5 Move Cyclicity and Generic Rhetorical Structure

As can be inferred from Table 4.5, except for the *Providing background information (RBI)* move, all the moves of the results sections recurred after being intervened by other moves. On average, *Presenting results (RRR)* recycled 17 times per article. *Commenting on specific results (RCR)* averaged three times, and *Summarizing results (RSR)* twice.

Table 4.6 below displays high frequency multi-move sequences widely observed in the methods sections. Move patterning was found to be very common in the texts. The most remarkable pattern was RRR-RCR which repeated 241 times in 75 texts, averaging more than three times per text. Another 2-move sequence was RBI-RRR, appearing 71 times in 63 texts, indicating a strong association between the two moves. A more complex pattern was RRR-RCR-RRR-RCR, with an average recurrence rate of three times per text in nearly half of the total number of results sections. Interestingly, with increased degree of complexity, the average recurrence rates of those patterns remained a constant three. However, this is not surprising because those patterns appeared mostly in the same texts. A more involved pattern actually embraced one less complex. In

other words, the phenomenon of a constant average recurrence rates of the patterns was

due to the fact that less complex patterns were repeatedly counted.

Table 4.6 Multi-move Sequences in the Results Section

Multi-move sequence	Frequency	Range	Probability	Rank
RRR RCR	241	75	0.15	1
RCR RRR	221	72	0.18	2
RRR RCR RRR	216	72	0.14	3
RCR RRR RCR	155	51	0.13	4
RRR RCR RRR RCR	154	51	0.10	5
RCR RRR RCR RRR	135	47	0.11	6
RRR RCR RRR RCR RRR	134	47	0.08	7
RBI RRR	71	63	0.19	8
RBI RRR RCR	36	35	0.10	9
RBI RRR RCR RRR	36	35	0.10	10

Again, as in the other sections, recurrences of move sequences suggest

that the unit of recycling was not merely individual moves but also sequences of

moves. Some recurring sequences were bound together to form a larger unit serving a

more complex communicative purpose.

The move codes files show that 60 out 120 results texts had *Providing*

background information (RBI) as the first move. In the absence of RBI, all but one

results sections began with an RRR.

As to the generic move structure of the results section, the results

presented above indicate an overall pattern of (RPR)-RRR-RCR-(RSR), with RRR

being obligatory, RCR conventional, and RBI and RSR optional.

In accordance with this overall generic structure (see Table 4.5), when

writing a results section, education research writers may begin with reiterations of

research objectives, data analysis procedures or descriptive statistics of the sample as a way of preparing the reader for results soon to be reported (RBI). Then, they report their findings (RRR) one by one, clearly and objectively articulating them in correspondence with their research questions (RRR2) and illustrating them with examples or excerpts (RRR3). However, before reporting their findings, they may remind the reader of whatever necessary for understanding the findings. Finally, they tend to briefly comment on the findings just presented (RCR), interpreting them, accounting for them, or comparing them with previous literature. Contingent on the circumstances, they may follow up with a recap of the major findings (RSR).

4.1.5 The Discussion Section

Table 4.7 displays the moves and steps observed in the discussion sections. Altogether, five moves were identified: *Providing background information (DBI)*, *Commenting on specific results (DCR)*, *Evaluating the study (DES)*, *Making deductions from the study (DDS)*, and *Summarizing the study (DSS)*. DCR was the only obligatory move and DDS is highly conventional and can be considered quasi-obligatory. The cut-off percentage of 60% establishes DBI and DES as conventional and DSS as optional. Furthermore, DCR, DES and DDS had steps but DBI and DSS had none. The following subsections present the findings in detail.

Table 4.7 Rhetorical Move Structure of the Discussion Section

Code	Move	f	n	%	S
DBI	**Providing background information**	**124**	**90**	**75**	**Con**
DCR	**Commenting on specific results**	**1560**	**120**	**100**	**Obl**
DCR1	Restating hypotheses or research questions	91	42	35	Opt
DCR2	Restating results	567	118	98.3	Con
DCR3	Interpreting results	321	101	84.2	Con
DCR4	Comparing results with literature	283	99	82.5	Con
DCR5	Accounting for results	223	89	74.2	Con
DCR6	Illustrating with examples or excerpts	35	25	20.8	Opt
DCR7	Recommending future research	48	29	24.2	Opt
DES	**Evaluating the study**	**205**	**95**	**79**	**Con**
DES1	Indicating significance	79	58	48.3	Opt
DES2	Acknowledging limitations	129	76	63.3	Con
DDS	**Making deductions from the study**	**370**	**115**	**95.4**	**Con**
DDS1	Drawing implications or making suggestions	205	102	85	Con
DDS2	Recommending future research	165	86	71.7	Con
DSS	**Summarizing the study**	**54**	**51**	**42.5**	**Opt**

Note. f = frequency of a move/step in all the RAs; n = number of RAs in which a move/step

appears; % = percentage of a move/step in all the RAs; S = Status; Obl = obligatory; Con =

conventional; Opt = optional

4.1.5.1 Providing Background Information

The statistics in Table 4.7 indicate that a *Providing background*

information (DBI) move was broadly distributed in the ERC articles, although its raw

frequency was relatively low (75%; f = 124; n = 90). This conventional move

provided whatever information necessary for understanding the discussion of the

results. Naturally, if present, it always occupied the initial place in the discussion

sections.

> 75) *This study is among the first randomized, controlled trials to*
> *examine different forms of professional development in early childhood and their*
> *impact on quality language and literacy practices. Results indicated that coaching*
> *was a more effective professional development form than coursework for improving*
> *the structural characteristics in classrooms. Differences among groups on these*

structural features were educationally meaningful, with effect sizes moderate to large in these areas of change. These improvements were maintained and, to some degree, enhanced 5 months after the intervention was over. ....(ESJ_09)

76) Sociocultural theories of learning, in which this study is rooted, strongly emphasize the influence of context on learning, behavior, and interactions. The context in which learning, or any activity, takes place can include elements such as the physical setting, demands on and opportunities for behaviors in that setting, individuals, and relationships among them. Sociocultural theory, which maintains that learning is constructed in interactions among individuals, also highlights the importance of discourse in processes of learning, a position which is substantiated by years of research on classroom discourse. This perspective on learning, as well as a relative lack of research on talk between teachers and students in a museum setting, led us to hypothesize that there might be differences in teacher-student discourse between these settings—differences that may also help illuminate some of the processes by which learning occurs on school trips. And indeed, differences were found. .... (RST_3)

In the discussion sections investigated, the backdrops that the writers provided for their discussions of results featured different topics, but mostly purposes, hypotheses, methods, or major outcomes. In Excerpt 75, the author recapitulates the research purpose, methodology, and principle results. Differently, in Excerpt 76, the author reiterates in relative detail the theoretical framework, the gap in previous research and the hypothesis, but makes no clear statement about the research outcomes.

4.1.5.2 Commenting on Specific Results

Table 4.7 reveals that both in terms of overall frequency and distribution (100%; f =1560; n = 120), *Commenting on specific results (DCR)* was by far the most important move in the discussion sections. This obligatory move communicated the authors' interpretations of specific results in relation to the general context and findings of previous research, the underlying reasons for the outcomes, and their

generalizability. This communicative purpose was realized by certain configurations

of 7 steps: *Restating hypotheses or research questions (DCR1)*, *Restating results*

(DCR2), *Interpreting results (DCR3)*, *Comparing results with literature (DCR4)*,

Accounting for results (DCR5), *Illustrating with examples or excerpts (DCR6)*, and

Recommending future research (DCR7). Details about these steps are presented in the

subsequent sub- subsections.

4.1.5.2.1 Restating Hypotheses or Research Questions

This step was observed in one third of the discussion texts where

it recycled more than twice on average (35%; f = 91; n = 42). Taking this optional step,

the writers reminded the reader of a specific hypothesis or research question or

anticipated outcome, as the excerpts below demonstrate.

> 77) *For our second research question, we asked, "Does a campus*
> *where students take greater advantage of those racial diversity-related opportunities*
> *have independent positive effects on students' learning and educational experiences?"*
> *..(AER_03)*

> 78) *In our first hypothesis we posited that the number of lectures*
> *students viewed online and attended in person contributed positively to course*
> *performance. ....(CAE_03)*

As in the results sections, a direct link was often observed

between a specific hypothesis or research question and a result being discussed.

Indeed, many discussion sections were organized under different sub-headings of

hypotheses and research questions or under sub-headings of different research

instruments. The clarity of this structure should facilitate the reader's understanding.

4.1.5.2.2 Restating Results

Restating Results (DCR2) was found in all but two discussion texts (98.3%, f = 567; n = 118), with an average recurrence rate as high as nearly six. The purpose in reiterating a specific result was to alleviate the reader's cognitive burden in following the author's thoughts about the result being commented on.

79) We found that when a video is shown two times, once with captioning and once without, the order of viewing has an effect on the subsequent recognition of vocabulary presented in the aural mode: learners presented with captions in the first viewing are better able to aurally recognize novel vocabulary than learners presented with captions in the second viewing...........................(LLT_04)

80) There were significant differences between the survey responses of the male students and those of their female classmates. In addition, the males in the Experimental group achieved a significantly higher mean score than the females in the post-intervention quiz. However, for the Control group, there were no significant differences between males and females in terms of survey responses and performance test scores. .... (JEE_05)

There was a marked difference between results reiterated in discussion sections and those in results sections. Whereas statistics were necessary in results sections, no statistics were given in the restatements of results in discussion sections, as demonstrated by the excerpts above. After all, the focus here was on the comments made thereafter.

It is noteworthy that, as in the following excerpt, a considerable number of DCR2 texts compared with previous literature at the same time. However, as the most-prominent-function move identification criterion dictates, such cases were assigned to a DCR2 step, regardless of secondary purposes.

81) While several smoking studies have demonstrated the impact of social appeals (Citation), in the current content analysis their presence was rare.........
... (HER_10)

4.1.5.2.3 Interpreting Results

The corpus revealed a wide use of the *Interpreting the results (DCR3)* step (84.2%; f = 321; n = 101), second only to DCR2 among the constituent steps of DCR. It was repeated an average of three times per text. As shown in the following excerpts, this conventional step went a step further from the objective finding and explained its meaning in non-technical terms comprehensible even to non-members of the discourse community.

82) This finding indicates Internet is an important tool for e-citizenship. ....(CAE_10)

83) In other words, the benefits of interacting with others of another race or ethnicity on racial- cultural engagement seem to be even more significant for students when their campus peers are interacting less frequently. ....(AER_03)

Again, like DCR2s, a few DCR3s were found to embed elements referencing previous literature.

4.1.5.2.4 Comparing Results with Literature

As Table 4.7 shows, as important as the results interpretation step (DCR3), *Comparing results with literature (DCR4)* was observed with a high frequency and a broad distribution (82.5%; f = 283; n = 99). On average, this step recurred nearly three times per text. This conventional step was similar to *Reviewing items of previous research (IET4)* in the introduction section in that both cited

previous literature. However, they differed significantly in communicative purpose.

Whereas IET4s were aimed at creating a space for the current study by uncovering a

gap or deficiency in past research, DCR4s were intended to justify findings or support

explanations by showing similarities or differences between them and those of past

literature. Of the excerpts below, one suggests disagreement and one consistency.

> *84) In contrast, when examining a large corpus of teacher-student classroom discourse, Nystrand et al. (2003) found that student questions did have an important role in promoting dialogic spells and open discussion, but other research reflects that such modifications also depend on teacher responses (Aguiar et al., in press). .... (RST_03)*

> *85) Thus, our findings are consistent with Desimone's (2002) insights from the literature on comprehensive school reform. Namely, the more specific, consistent, authoritative, powerful, and stable a policy is, the stronger its implementation will be. ....(ESJ_10)*

4.1.5.2.5 Accounting for Results

Table 4.7 reveals another conventional step, *Accounting for results*

(DCR5) (74.2%; f = 223; n = 89). It appeared more than twice per text on average.

Taking this step, the authors of those DCR5 texts gave further reasons for results,

more commonly results unexpected or inconsistent with those of past research.

> *86) This is probably due to the fact that these children were given a static model deprived from any motion information, in particular, information about movement direction. .... (LAI_01)*

> *87) Given the well-documented reciprocal relationship between phonological awareness and knowledge of grapheme-phoneme correspondences, it could be that the more severe phonological processing difficulties shown in this group stem from additional difficulties in mapping phonemes to graphemes.............(SSR_04)*

As shown in the above excerpts, authors exercised their power of reasoning on the basis of knowledge of real life situations or reliable findings of previous research.

4.1.5.2.6 Illustrating with Examples or Excerpts

A number of the authors of the ERC articles were also observed to take an optional *Illustrating with examples or excerpts (DCR6)* step when commenting on their results (20.8%; f = 35; n = 25). As the following excerpts show, they did so in order to clarify and support results or claims they made based on the results.

88) For instance, children who reported having an older brother who was involved in structured PAs or a younger sister who enjoyed going to the local park appeared to be influenced to do the same. Conversely, children who had an older brother who spent considerable amounts of time playing E/PCG were also influenced to do the same. ....(AER_02)

89) Consider the use of hedging in cautious language for example. L1 Chinese learners of L2 English in the current study are found to show some control of this feature in their academic writing, but do not demonstrate it as diversely and robustly as native writers do...(LLT_05)

The DCR6 step shared the same purpose with the *Illustrating with examples or excerpts (RRR3)* step of the *Reporting specific results* (RRR) move in the results section. However, it was much less frequently observed than RRR3. With results already illustrated with examples, there was not much of a need to do the same. Needless to say, the main focus here was on the commentary itself.

4.1.5.2.7 Recommending Future Research

Another optional step *Recommending future research (DCR7) was*

observed when commenting on specific results (24.2%; f = 48; n = 29). As the excerpts

below show, when commenting on a specific result, not a few authors suggested

follow-up research closely related to the result. Recommendations might take the form

of a direct question as in Excerpt 90, or a need statement as in Excerpt 91.

> *90) However, might the lure be not the status of the major, but its*
> *structure? ....*
> *.. (JEE_07)*

> *91) Further longitudinal research that looks at the longer-term*
> *impact of science identity development over the schooling years is critically needed....*
> *(RST_08)*

Given that later on in the same section, recommendations about

future research were to be made too, it seems that these authors were very much

aware that that part of the section was reserved for recommendations based on the

entire study rather than on a specific result.

4.1.5.3 Evaluating the Study

Table 4.7 shows that *Evaluating the study (DES)* was an conventional move

with a high frequency of observations and a wide distribution across the discussion

sections (79%; f = 205; n = 95). This move allowed the authors to appraise the entire

study in terms of worth, significance, weaknesses or limitations. This communicative

purpose was achieved by both or either of two steps: *Indicating significance (DES1) and*

Acknowledging limitations (DES2). Details of the two steps are as follows.

4.1.5.3.1 Indicating Significance

Being optional, *Indicating significance (DES1)*, was employed in approximately half of the discussion sections (48.3%; f = 79; n = 58). As the following excerpts show, the authors might highlight the importance of a study either in terms of its contribution to the current knowledge base or in terms of its methodological innovation deemed beneficial to future research.

92) ...the ... study potentially does add significantly to current knowledge ... Our study is one of the first to examine the relationship between AA of FV ... This cross-sectional study offers preliminary evidence .. .. (Abbreviated)(HER_01)

93) The scales that we have developed are of particular importance for future quantitative research on ICT integration in schools. As there are no available scales to measure ICT school and teacher level conditions, our study is useful for empirical progress in this field. ....(CAE_06)

4.1.5.3.2 Acknowledging Limitations

Nearly two thirds of the authors of the ERC articles were found to take an *Acknowledging limitations (DES2)* step in evaluating their research (63.3%; f = 129; n = 76). By this step, the researchers honestly identified limitations of their study and explained in what ways their study was limited. As the excerpts below demonstrate, the authors reflected on the nature of their study with respect to research design, sampling, instruments, setting, and other factors that might constitute a threat to the generalizability of their findings.

94) The primary limitation of this study is that it utilizes a non-experimental design. As such, no claims of causality can be made. This study is further limited by the HLE itself, as these findings may not generalize to other computer-based learning environments. In addition, the sample size for this study was

relatively small, which may be a reason why certain relations (i.e., planning SRL processes and conceptual knowledge posttest scores) failed to achieve statistical significance. Likewise, the sample was from a single high school, and more research is needed with larger and more diverse samples before definitive claims can be made regarding how historical thinking can be fostered using HLEs. ....(CAE_01)

95) The results of this study must be considered in light of its limitations. An important limitation is the low number of female participants (15 out of 80, although there were no between-group gender differences). Another limitation is that the instrument ... A further limitation is the manner in which the participants were grouped... One threat to generalizability is the treatment media... The results of this study are also limited in their ability to generalize to the English as a Foreign Language population at large, or to other topics or aspects of second language acquisition.. (Abbreviated) (LLT_07)

4.1.5.4 Making Deductions from the Study

Table 4.7 shows that all but five discussion sections made a *Making deductions from the study (DDS)* move (95.4%; f = 370; n = 115). This move, mostly observed towards the end of a discussion section, recurred three times per text on average. By this move, the authors synthesized individual results and made logical inferences with respect to what the findings meant and what important directions would be for future research. Typically, this communicative purpose was jointly realized by two steps, *Drawing implications or making suggestions (DDS1)* and *Recommending future research (DDS2)*, which are described in the following sub- subsections.

4.1.5.4.1 Drawing Implications or Making Suggestions

Table 4.7 indicates that *Drawing implications or making suggestions (DDS1)* was a very important conventional step (85%; f = 205; n = 102). In the ERC articles, this step was mostly placed under general or specific

sub-headings such as "Implications," "Pedagogical Implications," "Implications for Instruction," and "Theoretical Implications." The majority of the authors employed this step twice on average. They pointed out mostly practical, sometimes theoretical, and sometimes methodological implications that their study might have.

96) The reported study has important practical implications for engineering education. ... we found that .. In contrast, college-level textbooks mostly ... To our knowledge, the research base for ... is limited. The findings of the present study, although preliminary, suggest that precollege engineering instruction should also focus on ... Because pre-college students .., this practical implication is developmentally appropriate for this age. .... (JEE_04)

97) In conclusion, results of the present study carry theoretical implications. They support the application of paired associate learning concepts as a means of understanding how mnemonics contribute to the process of learning alphabet letters. Findings show the value of visual imagery in facilitating letter learning as predicted by dual coding theory (Citation). Children's performance on the transfer tasks supports expectations based on Ehri's (Citation) phase theory regarding the contribution of letter knowledge along with phoneme segmentation in helping beginners remember how to read and spell words..(SSR_02)

In the above excerpts, the authors extend beyond the objective findings and make inferences in presenting their study's theoretical, practical, and methodological implications. Yet, their inferences remain within the restraints of results, which are briefly reviewed once again.

4.1.5.4.2 Recommending Future Research

Recommending future research (DDS2) (71.7%; f = 165; n = 86) was observed in well over half of the discussion sections of the ERC articles, with an average recurrence rate of two per text. The writers suggested further research most

desirable to overcome the limitations already articulated. Specifically, the step focused on unanswered questions and aspects of the research that should be improved to enhance validity and generalizability.

> *98)   Future research may utilize other sources or measures of CEP to replicate the findings presented here.*...*(LAE_09)*

> *99) Clearly, more research is needed that examines ... with diverse samples. Causal relations between these phenomena can be examined .... Likewise, the HLE used in this study was static, but ... by using adaptive computer environments. Finally, studies are needed to determine whether the skills learned using HLEs transfer to other learning tasks not involving computers.*.........*(Abbreviated) (CAE_01)*

Both excerpts above suggest specific new research agendas aimed at addressing the limitations of the study. In just a few words, Excerpt 97 recommends replicating the study using another instrument. In contrast, Excerpt 98, which is abbreviated, is quite lengthy and much more comprehensive in scope. It attempts to address several limitations at once, revolving around enhancing the generalizability of results — by using different samples, switching to an experimental design, applying the curriculum to a different environment, and changing the instructional mode.

It should be noted that this step (DDS2) was similar in purpose to the *Recommending future research* (DCR7) step of the *Commenting on specific results* (DCR) move. Nevertheless, the two were different in the basis on which recommendations were advanced. While unsolved questions relevant to specific results gave rise to DCR7s, it was in light of the entire study that DDS2s recommended further studies, as exemplified by Excerpts 98 and 99.

4.1.5.5 Summarizing the Study

A *Summarizing the study (DSS)* move (42.5%; f = 54; n = 51) was observed in one third of the discussion sections, hence an optional move. This move, if present, concluded the discussion sections, or in 14 articles, appeared after the discussion section under a separate sub-heading, e.g., "Conclusion" or "Summary."

100) We have argued that while ..., what appears to have changed is ... Business schools still educate future business leaders ..., but this model is rarely seen as problematic by MBA students. Yet it... Indeed it has been shown that ... The idea ... was generally discounted. In further research we will ... For instance, we intend to analyze. ....(Abbreviated)(LAE_02)

101) In conclusion, a combination of learner interaction via synchronous text-based CMC combined with postchat corrective feedback seems to be a CALL learning strategy with great potential for improving both L2 communicative competence and accuracy. It may offer "the best of both worlds" combining the benefits of negotiation of meaning and resultant implicit corrective feedback in the synchronous mode with explicit corrective feedback in the asynchronous mode. This learning strategy certainly merits further investigation. ....(LLT_01)

As illustrated by the above excerpts, authors took this last chance to restate their major findings in answer to their research questions or hypotheses and recommend further research, finalizing the great cycle from research questions, to data collection and analysis, to results, and back to research questions.

4.1.5.6 Move Cyclicity and Generic Rhetorical Structure

Table 4.7 suggests that move recycling was common in the discussion sections. Of the five moves, *Commenting on specific results (DCR)* was the most cyclic, its recurrence rates averaging 13 times per text. *Making deductions from the study (DDS)* recurred three times per article. Similarly, *Evaluating the study (DES)*

averaged twice per article. However, the other two moves, *Providing background information (DBI)* and *Summarizing the study (DSS)*, seldom recycled in the texts.

As revealed by Table 4.8, move patterning is quite obvious in the discussion sections. Each of the 2-move patterns on top of the list, namely, DCR-DDS, DES-DDS, DDS-DCR and DBI-DCR, repeated approximately twice on average in the majority of the articles. The 3-move patterns, DCR-DDS-DCR, DDS-DES-DDS, and DDS-DCR-DDS, also recurred with high frequencies in about one thirds of the articles. Surprisingly, even the 4-move pattern, DCR-DDS-DCR-DDS, recurred nearly twice in one third of the articles. This rather involved 4-move pattern comprised the 2-move pattern DCR-DDS repeated twice, indicating structural complexity of the discussion section.

Table 4.8 Multi-move Sequences in the Discussion Section

Multi-move sequence	Frequency	Range	Probability	Rank
DCR DDS	153	85	0.13	1
DES DDS	112	68	0.20	2
DDS DES	92	57	0.10	3
DDS DCR	88	52	0.10	4
DBI DCR	82	82	0.15	5
DCR DDS DCR	81	47	0.07	6
DDS DES DDS	69	41	0.08	7
DDS DCR DDS	66	41	0.07	8
DCR DES	65	55	0.06	9
DCR DDS DCR DDS	63	38	0.06	10
DBI DCR DDS	51	51	0.09	11
DCR DES DDS	43	43	0.04	12
DCR DDS DES	39	39	0.03	13

Again, high recurrence rates of multi-move sequences provide further evidence that in the discussion sections of the ERC articles, move recycling involved units larger than moves as well.

Additionally, the move codes files suggest that the majority of the discussion sections began with a *Providing background information (DBI)* move (87 out of 120 articles). In cases where a DBI move was absent (28 out of 120 articles), most texts had an initial *Commenting on results (DCR)* move. The remaining five texts began with *Evaluating the study (DES)* or *Summarizing the study (DSS)*.

Based on the frequencies of the multi-move sequences and the relative positions of the moves, an optimal generic move structure was constructed: DBI-DCR-DDS-DES-(DSS), with DCR being obligatory; DBI, DES, and DDS being conventional; and DSS optional (see Table 4.7). Following this model, education writers may open a discussion section by providing the reader information necessary to enable their understanding of the discussion (DBI). Then they invariably comment on specific results (DCR), restating the results (DCR2), interpreting them (DCR3), comparing them with previous research (DCR4) and giving explanations for them (DCR5). Sometimes they need to initiate their commentary with a restatement of their hypotheses or research questions (DCR1) or end their commentary calling for further research to tackle whatever puzzles unresolved (DCR7). This commentary move can repeat as many times as there are results needing to be commented on. After the commentary, they tend to evaluate the entire study (DES) with respect to its significance (DES1) and limitations (DES2). Subsequently, they almost always make deductions from the study (DDS), highlighting its theoretical, practical or

methodological implications (DDS1) and suggesting future research agendas aimed at overcoming the inherent limitations (DDS2). Finally, if necessary, they may end their discussion with a summary of the main points of the study (DSS).

4.2 Discussion

Overall, the results indicate that education research writers use similar rhetorical moves predicted by the initial coding scheme and that when structuring the four sections of their articles they generally follow the same schematic structures on the initial coding scheme based on previous studies. However, despite general affinity with the move schemes, they may also use some other rhetorical moves or steps not accounted for by the initial coding scheme. These deviations may be considered distinct characteristics of the education RA genre.

The following subsections attempt to interpret and account for the results in light of previous literature and the realities of the discourse community of education researchers.

4.2.1 The Introduction Section

Except for the occasional use of the *Describing the conceptual model (IDM)* move, the introduction sections were conventional insofar as they generally followed the move structure proposed by the CARS model (Swales, 1990). Nevertheless, some deviations from the model were noted at both the move and step levels.

The most interesting deviation was perhaps the *Describing the conceptual model (IDM)* move that was quite common in the corpus. This move was not independent because it was based on a thorough review of previous theories and literature. However, given its distinct communicative purpose and the way it stood out, this study considers it a move in its own right. Although it has never been documented previously, it might be hasty to jump to the conclusion that it was nonexistent in the corpora used in previous studies. It might be equally hasty to rule out its existence in the remainder of the articles in the ERC. Quality research in any field is always carried out within a conceptual framework based on one or a set of relevant theories. Articles published in prestigious peer-reviewed international journals can be assumed to be of high quality. Therefore, they should not lack a conceptual framework. The majority of the articles in the ERC corpus did include a conceptual framework, but it was often interwoven into literature being reviewed, hence without a definitive shape. This may explain the move's absence in previous descriptions of the RA rhetorical structure. Indeed, as such, it might have been overshadowed by the *Reviewing items of previous research* (IET4) step. The present study discriminates descriptions of conceptual frameworks intertwined with literature and those that stand alone. The purpose of this distinction is solely pedagogical.

Step-level differences involving all moves in the CARS model are equally revealing about the education RA genre. Consistent with the CARS model (Swales,

1990), the first move *Establishing a territory* (IET) was found obligatory in the ERC corpus, which was realized by several steps. However, new steps were observed and some steps were found to differ in some respects. Although the degree of dominance of the *Claiming centrality* (IET2) step was very close to that of the same step found in NS educational psychology RAs (Loi, 2010), it was much more frequently employed in the ERC articles than in physics (Swales, 1981) and educational psychology RAs (Swales & Najjar, 1987). One possible explanation for the difference is that the need to assert the importance of their research is more keenly felt by education researchers competing for recognition in the broader field of education than those working in narrower disciplines such as physics.

The results also revealed that the *Reviewing items of previous research (IET4)* step played a very important role in territory establishment both in terms of frequency and distribution. While its high concentration in the introduction sections investigated supports Antony's (1999) finding about software engineering RAs, it contradicts with previous authors who recommended only a brief review of past literature in introductions (Brusaw, Alred, & W. E. Oliu, 1993; Day, 1977; Harmon, 1989). The contrast between this finding and Anthony's and those authors' recommendations probably reflect the evolutionary nature of the research article genre. Whereas within an emergent discipline, there is not much previous work to relate to, within an established discipline that has accumulated a great body of literature over the years, a writer normally needs to situate his or her study in the context of a large number previous studies.

As for the *Defining terms and concepts (IET5)* step which is absent from Swales' CARS model (1990), its highly optional status in the ERC articles is in keeping with Anthony's (1999) and Loi's (2010) observations of its the extensive use in software engineering and educational psychology. Due to its prominence, Anthony even considered it a separate move after territory establishment. There might be a greater need to clarify certain blurry concepts or to operationalize certain variables in fields of science and technology that are still emerging and therefore lack standardized terminology, or in social sciences or interdisciplinary fields like education research that are well-established but have ambiguous or overlapping concepts due to multiple theories.

Another anomaly of territory establishment was a *Presenting a vignette (IET1)* step found at the onset of a few introductions. None of the previous literature has mentioned IET1 except for Swales (1990) who noted an instance of IET1 in one out of four articles in an issue of *Research in the Teaching of English*. Having difficulty in assigning it to any of the moves or steps in his CARS model, Swales assumed that the "anecdote," as he called it, functioned simply as an opening attention-catcher. Given that the vignettes found in the ERC all fulfilled the purpose of introducing the reader to a research topic and were always adjacent to an *Establishing a niche (IEN)* move, categorizing them as an *Establishing a territory (IET)* move seems appropriate. It is noteworthy that the five IET1s were concentrated in two journals only. This step could well be journal-specific. Or, perhaps, it reflects possible disciplinary variation.

Step-level differences were also observed with respect to the *Establishing a niche (IEN)* move. In particular, rare cases of a *Postulating hypotheses (IEN1E)* sub-step were observed mostly concentrated in one particular journal, *Sociology of Education*. It seems logical to claim that hypothetical statements are an idiosyncratic trait of this particular journal. As stated in its policy, this journal focuses on research that examines how social institutions and individuals' experiences affect educational processes and social development. Such research is often in the form of experimental or correlation studies that involve hypothesis testing. This sub-step is undocumented in previous research. One explanation regarding its rare use might be that education researchers tend not to posit hypothesis in a formal manner as do natural scientists who rely on experiments to test hypotheses. Kpolovie (2011) may have gone too far when he said that it was the absence of appropriate hypotheses that resulted in the inferior quality of the great majority of educational journal articles.

A *Providing positive justifications (IEN2)* step was observed frequently used in combination with gap identification. This finding lends support to Samraj (2002) who found writers of wildlife behavior RAs occasionally stressed the need for their research when establishing a research space. According to Samraj, this step is advantageous over the other niche establishing steps because it constitutes no threat to other researchers' face. In his revised CARS model, Swales (2004) included it as an optional step. This finding provides additional evidence that a research space can alternatively be established by emphasizing the need for a particular piece of research.

Turning to the *Presenting the present research* (IPR) move, interesting differences were indicated at the step level too. In the ERC corpus, two types of IPR1s were found in the corpus: general and specific. General IPR1s, always placed towards the end of the introduction section, previewed the study in summary form covering several aspects of it such as purpose, research design, hypotheses, and research questions. In contrast, specific IPR1s tended to mention one particular aspect of the current study once a specific research gap was identified before. While previous investigations invariably reported on the general IPR1 step, claiming that it always appeared towards the end of the introduction section (e.g., Pho, 2008a; Samraj, 2002; Swales, 1990; Y. Zhang & Hu, 2010), this study uniquely identified a large number of specific IPR1s. Since a number of gaps or deficiencies were often identified at different stages of a systematic literature review, writers might find it both convenient and necessary to announce each aspect of their study immediately after each gap was pinpointed.

The *Announcing principal findings* (IPR2) step was found to be thinly distributed in the ERC corpus. This finding is in line with previous findings about RAs in education research (Swales & Najjar, 1987) and applied linguistics (Pho, 2008a). Yet, it is in sharp contrast with findings on science and technology such as biochemistry (Kanoksilapatham, 2007), civil engineering (Kanoksilapatham, 2011), agriculture (Del Saz-Rubio, 2011) and physics (Swales & Najjar, 1987) where major findings were much more frequently previewed. This finding seems to provide

corroborating evidence that education researchers refrain from announcing their findings early in introductions. This common discursive practice is perhaps related to the need felt by education researchers to avoid repetition. After all, since major findings have already been presented in the abstract preceding the main body, doing away with an announcement of principal findings in the introduction will save some valuable space in an article that typically runs 30-40 pages.

The *Indicating RA structure (IPR3)* step was present in a quarter of the ERC articles. The finding is inconsistent with those of previous studies. While it was observed in at least half of engineering (Cooper, 1985) and applied linguistics (Sheldon, 2011) RAs, it was present in less than 1% of biology and wildlife behavior RA introductions (Samraj, 2002). Can this marked inconsistency be explained in terms of disciplinary variation? The aforementioned studies were all based on corpora of a limited size. No substantial evidence suggests either way. Alternatively, awareness of the article length allowed might be a determinant. Perhaps longer articles are more likely to contain this step.

The *Claiming significance of the research (IPR4)* step was found in a quarter of the ERC articles. According to Swales (2004), IPR4 is probable in certain fields but unlikely in others. This step has been observed in agricultural science, educational psychology and civil engineering RAs where its distribution ranges from 10% to 38% (Crookes, 1986; Del Saz-Rubio, 2011; Kanoksilapatham, 2011; Loi & Evans, 2010; Ozturk, 2007). With a figure closer to the figure in educational psychology, the

present study indicates that positively evaluating the present research can be an option in the field of education research.

Interestingly, a *Verifying compliance with ethical standards (IPR5)* step was observed in a single introduction section. Since an identical step was observed much more frequently in the methods section, an explanation is offered when discussing findings about the methods section.

Documented only in Kanoksilapatham (2015), a *Describing research context (IPR6)* step was also found optional in the introductions. This finding suggests that education research writers can either provide contextual information in the introduction as well as in the methods section.

With respect to the behavior of moves in relation to other moves, the results indicate that with the exception of *Describing the conceptual model(IDM)*, all the moves were cyclic and they generally followed an IET-IEN-IPR scheme. Moreover, recycling did not just involve individual moves. Cycles of two or three moves were readily discernable. The cyclic nature accords with previous studies of RAs in various fields (Crookes, 1986; Del Saz-Rubio, 2011; Loi & Evans, 2010; Ozturk, 2007). This marked cyclicity resulted from the great lengths of the education RAs in the corpus. As Crookes (1986) and Swales (1990) rightly observed, move cyclicity correlates with length of text.

4.2.2 The Methods Section

In comparison with the other sections of the ERC articles, the methods sections investigated revealed a limited number of ways in which they seemed at odds with previous descriptions.

The results suggest that all ERC articles used the *Describing data collection procedures (MDC)* move and the majority employed the *Describing data analysis procedures (MDA)* move as predicted by the initial coding scheme. However, half of the articles began the methods section with an additional *Describing research design (MRD)* move, which is undocumented in previous research. This indicates a unique feature of education research articles.

Regarding the *Describing data collection procedures (MDC)* move, the most remarkable is the *Describing research context (MDC1)* step present in over one third of the ERC articles. This step was almost completely absent from the findings of previous genre studies. One interesting study (Huang & He, 2011) identified texts segments describing the location/source of the sample, but categorized them as a constituent step of the *Describing the sample/population* move. Obviously, their *Describing the location/source of the sample* step was narrower in scope than the *Describing research context (MDC1)* step in the present study. As will be discussed in detail later, this discrepancy points to variation linked to different research paradigms.

A *Justifying data collection procedures (MDC5)* step was also observed in a few ERC articles. It is somewhat difficult to compare this finding with those of previous research. In her comparative study of applied linguistics and education technology RAs, Pho (2008a) found it present in the former discipline but not in the latter. Lim (2006) observed it in management RAs, but the justification step in Lim's study was narrower in scope because it only concerned sampling, as indicated by its two sub-steps: highlighting advantages of using the sample and showing representativity of the sample. Perhaps, this suggests another idiosyncrasy of the move structure of education RAs. As indicated by Swales' clipped-elaborated cline of variations in methods sections (2004), elaborated texts tend to use more detailed justification steps than clipped texts. The average length of the articles in the ERC, which is 8,131 words per article, suggests that education RAs belong to the elaborated category.

Intriguingly, a *Verifying compliance with ethical standards (MDC6)* was found in a considerable number of methods sections as opposed to only one observation of an identical step of the *Presenting present research (IPR)* move in the introduction section. This finding corroborates that of Kanoksilapatham (2015) who found the step quite common in biochemical engineering RAs. It suggests that in education research as in biochemical engineering, among others, human subject protection is of supreme importance. Precautions must be taken to ensure that harms and risks are minimized, benefits maximized, and human dignity, privacy and autonomy respected. Normally,

when submitting a manuscript, authors have already included a verification statement in the cover letter as required by the APA style manual (American Psychological Association, 2010). To find out why such verification statements were present in some articles but absent in others, the researcher examined the authors' biographies and affiliation information. It was found that of all 17 articles with a verification statement, nine were authored by researchers affiliated with U.S. institutions and the other eight by researchers from eight different countries including the UK, Canada, Australia, Singapore, France, Belgium, Turkey and Greece. Interestingly, nine of the articles were concentrated in the journal *Health Education Research (HER)*, another four were published in the *Journal of Engineering Education (JEE)* and the rest in four other journals. This finding appears to suggest variations among journals, but we would be more confident to link the finding with a strong desire to get a manuscript accepted by the review panel and the readership as well. Whether it is required by the journals or not, as long the study is potentially threatening to human subjects, it will be wiser to play safe by including a statement to indicate conformity with established ethical requirements.

Concerning the *Describing data analysis procedures (MDA)* move, the present study found that a great majority of the ERC articles employed this conventional move, in support of Pho's (2008a) findings about RAs in applied linguistics and education technology. Nevertheless, whereas Pho's data analysis move did not have constituent steps, the complexity of the MDA moves in the ERC corpus required three

different step labels to categorize more specific communicative functions. These steps were *Recounting data analysis procedures* (MDA1), *Justifying data analysis procedures (MDA2)*, and *Establishing inter-coder reliability (MDA3)*. Both the addition of these steps itself and the nuanced step-by-step descriptions of the data analysis procedures suggest a keenly felt need of the authors to provide adequate details. For example, although found in only a small portion of the ERC corpus, MDA3 was typical of articles of qualitative research in which coding is frequently performed on qualitative data like transcripts of interviews. In fact, all the 26 articles containing an MDA3 step in the ERC corpus were qualitative in nature. This is not surprising since it is a conventional practice to involve multiple coders or raters so as to reduce the degree of subjectivity of results. It should be noted that the data analysis move was absent in nearly one-third of the methods sections in the ERC corpus. Nevertheless, this by no means indicates that the authors did not describe their data analysis procedures. As findings on the results section suggest, some writers postponed describing data analysis till the results section. This is particularly true of quantitative studies that involved established, standard statistical procedures.

With respect to move cyclicity, the methods sections in the ERC articles were found to be the least active in recycling. This is because authors, irrespective of disciplines, do not tend to juggle describing data collection procedures and data analysis procedures as these activities naturally occur one after the other.

In sum, a marked characteristic of the methods sections of the education RAs, particularly of the data collection and analysis moves, is the thoroughness of descriptions and the high degree of attention to detail accomplished by means of multiple steps. For certain, this is related to the fact that a considerable proportion of the ERC articles were qualitative. Qualitative research, largely concerned with understanding complex issues, embraces an interpretivist research paradigm and relies on such methods as interview, observation, ethnography, content analysis, etc. which are highly context sensitive and flexible. In the conduct of qualitative research, researchers often rely on an audit trail that documents in great detail all aspects of the research — context, sampling, roles of different research team members, roles of different data sources, data analysis procedures, etc. With regards to reporting data analysis procedures, Malterud (2001) stated that "the reader needs to know the principles and choices underlying pattern recognition and category foundation" (p. 486). Lincoln and Guba (1985) emphasized the need for "thick" descriptions of the context, sample and procedures of data collection and analysis. Detailed information can help the reader evaluate the extent to which the research findings are transferable to other times, settings, situations, and populations. For example, it would be difficult for the reader to understand research results without sufficient information about a broad context of the research setting in terms of the geographical location of the research site and important demographics of the general population such as age, sex, ethnicity, socio-economic status, educational level, political stance, etc.

4.2.3 The Results Section

Largely in congruence with the move scheme for the results section proposed by Pho (2008a), the results show that four moves were employed in the results sections of the ERC articles, including the obligatory *Reporting specific results (RRR)* move, the conventional *Commenting on specific results (RCR)* move, and the two optional moves of *Providing background information (RBI) Summarizing results (RSR).*

With regards to the *Providing background information (RBI)* move, the present study found it employed in nearly half of the ERC articles, suggesting that education researchers are more than likely to inform the reader once again of the general research purpose, the data analysis procedures, or descriptive statistics about the sample, thus paving the way for the reader in understanding the results. The presence of this move explains the absence of the *Describing data analysis (MDA)* move in the methods sections of a number articles that employed standard, established statistical data analysis procedures. The authors might feel more convenient to briefly mention them immediately before presenting results in the results section. This presentation technique shortens the psychological distance between results and the data analysis methods used to obtain them thus easing the reader's cognitive burden in assessing the results. Regarding this preparatory move, the rather limited previous research seems to divide over its use. On the one hand, no preparatory move was found in some disciplines. For example, Kanoksilapatham (2007, 2011) neither found it in

biochemistry nor in civil engineering. On the other hand, this move was quite common in applied linguistics and education technology (Lim, 2010; Pho, 2008a; Yang & Allison, 2003). Moreover, similar pre-results elements were observed in sociology, computer science, and medicine, whose communicative functions were similar although they were categorized as *metatextual*, *structure of section*, and *procedural* (Brett, 1994; Posteguillo, 1999; I. A. Williams, 1999). Kanoksilapatham (2015) recently found a similar pre-results move very common in three engineering sub-disciplines, although she termed it *Summarizing procedures*.

While a general preparatory move was identified in a number of previous studies, this is the first time that a *Preparing for specific results presentation* (RRR2) step has been identified. Its amazingly high frequency and wide distribution in the ERC corpus suggests that it is a very common practice at least for education researchers to provide information essential for assessment of specific results.

The extensive use of an *Illustrating with examples or excerpts (RRR3)* step in the ERC articles indicates another feature unique to education research articles of a qualitative nature. Taking the entire corpus as a whole, the RRR3 step was optional. However, since this step was only observed in qualitative articles, it is safe to claim that this step is conventional for qualitative papers. It is noteworthy that none of the previous studies that the researcher knows of has observed this step in the RA results section. This may well be due to the exclusion of qualitative research articles in past

studies of the RA genre. To date, genre analysts of the research article have largely

avoided qualitative papers, explicitly stating that their corpora consist of experimental

or empirical research articles. However, while it may be sensible to control variance

associated with different methodologies by excluding qualitative papers, articles

based on qualitative research certainly deserve investigation. Moreover, qualitative

research is empirical insofar as it is based on empirical data.

The *Commenting on specific results* (RCR) move was found optional in the

ERC corpus. This finding comes as no surprise because commenting on results in the

results section is not unusual in a broad range of fields such as computer science,

medicine, biochemistry, civil engineering, sociology, education technology, education

and applied linguistics (Basturkmen, 2009; Brett, 1994; Kanoksilapatham, 2003, 2007;

Lim, 2010; Pho, 2008a; Posteguillo, 1999; I. A. Williams, 1999; Yang & Allison,

2003). Conventional as it was in the results sections, the importance of the

commentary move was by no means comparable to its counterpart in the discussion

section. While the ratio of the total frequency of the move to the number of articles

where it appeared was two for the results section commentary move, the figure was

five for its counterpart in the discussion section. Unlike discussion sections where

different steps of the commentary move typically co-occurred in one single text,

results sections seldom had such comprehensiveness, although, as a whole, they

contained similar chunks of text that could be further classified as was the discussion

commentary move. This explains why, in the present study, the results section commentary move is not broken down into constituent steps. With respect to importance, Yang and Allison (2003) observed that the commentary move in results section plays a significantly less important role than its counterpart in the discussion section.

Given the opportunities available in the discussion section, what motivates commentary on results early in the results section is still empirically unknown. In contemplating why biochemistry results sections contain commentary elements, Swales (2004) quoted a botanist informant as saying "Why not preempt potential puzzlement and criticism by providing explanations and justifications at the earliest possible time?" The botanist's words and therefore Swales' tentative opinion might sound plausible. Yet, could this also reflect a dilemma that researchers face — whether to adhere to the traditional IMRD organization as prescribed by most publication guidelines or follow the trend towards a combined results and discussion section? The authors might have made a compromise between the rigidness of the conventional structure and the convenience afforded by an integrated results and discussion section. Along with the absence of constituent steps of the commentary move, the absence of limitations, implications and significance moves in the results sections may be proof that the authors steered a middle course.

Finally, an optional *Summarizing resulting (RSR)* move found in the ERC results sections demonstrate that education writers may end their results sections with a recap of their major results. Although this move was absent from the results section move scheme in the initial coding scheme based on Pho (2008a), it has been identified by a number of researchers, such as Posteguillo (1999), Lim (2010), Yang (2003), and Kanoksilapatham (2005). Except for Kanoksilapatham (2005) who considered it a step of the commentary move, all of them regarded it as a separate move. A subsequent examination of the articles containing this move revealed that the number of specific results was a major factor in determining whether to include an RSR move or not. The greater the number, the more likely the presence of this move.

When it comes to move recycling patterns, the results show that the RRR-RCR pattern tended to recycle three times per text on average. The complexity of education research apparently gave rise to this high frequency cycle. Evidence suggests that the number of times the cycle recurred positively correlated with the number of research questions, variables and data collection instruments that were involved in a study.

4.2.4 The Discussion Section

The authors of the ERC articles were found to make five moves in discussing their findings: *Providing background information (DBI)*, *Commenting on specific results (DCR)*, *Evaluating the study (DES)*, *Making deductions from the study (DDS)*, and *Summarizing the study (DSS)*. This move structure generally resembles those

proposed by other researchers for a range of disciplines (Amirian et al., 2008; Amnuai & Wannaruk, 2013b; Hopkins & Dudley-Evans, 1988; Kanoksilapatham, 2005; Peacock, 2002; Pho, 2008a; Swales, 1990; I. A. Williams, 2005; Yang & Allison, 2003). Despite that, minor differences existed mostly at the step level.

The results show that the *Providing background information (DBI)* move was conventional, being present in nearly two thirds of the ERC articles. This indicates that education researchers generally prefer to start their discussion of results with brief restatements of overall objectives, methods, or principal findings to ensure that the reader understands the logic behind the points they are about to make. Basturkmen (2009) states that providing background information reflects a special consideration of the writer for readers who only read the abstract and the discussion section. Previous investigations into the discussion section move structure in a wide variety of disciplines were not unanimous about the move's importance. In statistical terms, some studies indicated that this move played a very significant role with occurrence rates ranging from 60% to 90% in the texts investigated (e.g., Basturkmen, 2012; Kanoksilapatham, 2007; Yang & Allison, 2003). Others found the same move optional by the standard of the present study (e.g., Amirian et al., 2008; Amnuai & Wannaruk, 2013b; Peacock, 2002; Pho, 2008a; Posteguillo, 1999). This move was even absent in Nwogu's (1997) medical RA corpus. It would be premature to ascribe the variation to disciplinary differences because of the confusion that the results of the studies cause.

Peacock's (2008) corpus consisted of RA discussion sections in seven disciplines. Even worse, results of different studies of the same disciplines were conflicting. For instance, in applied linguistics discussion sections, the background information step was found conventional in one study (Yang & Allison, 2003) but optional in a few others (Amirian et al., 2008; Amnuai & Wannaruk, 2013b; Pho, 2008a). Yet, suffice it to say that the background information move is typical of education research articles.

The *Commenting on specific results (DCR)* move was found universal in the ERC articles. This move communicates the authors' interpretations of specific results in relation to the general context and findings of previous research, the underlying reasons for the outcomes, and the generalizability of the findings. The results show that this move had seven possible steps: *Restating hypotheses or research questions (DCR1)*, *Restating results (DCR2)*, *Interpreting results (DCR3)*, *Comparing results with literature (DCR4)*, *Accounting for results (DCR5)*, *Illustrating with examples or excerpts (DCR6)*, and *Recommending future research (DCR7)*.

The authors of the ERC articles, before actually making any comments, almost always restated a result or a group of related results (DCR2) and sometimes the research question or the hypothesis corresponding to the results (DCR1). Together, these two steps performed a similar function as the *Providing background information (DBI)* move, that is, to prepare the reader for understanding what followed. Yet, they were markedly different in scope. The background information move, expressed in one or more

full-length paragraphs, was general in that it prepared the reader for the whole discussion section. In contrast, the two restatement steps were only part of a paragraph functioning to prepare the reader for commentary on only one or a set of related specific results. This finding corroborates that of Kanopsilapatham (2005), who noted a similar distinction in biochemistry RAs between a general background information move (labeled *Contextualizing the study*) and two restatement steps under an umbrella commentary move (termed somewhat misleadingly *Consolidating results*), one restating purposes, research questions, hypotheses, and procedures; the other restating specific results (labeled *Stating selected results*). Other studies seem to obscure the general vs. specific distinction (e.g., Amnuai & Wannaruk, 2013b; Peacock, 2002; Yang & Allison, 2003). A typical example is Basturkmen's (2009) focused study on the commentary move. She reported three moves preceding the central commentary move, which she respectively termed *Background information*, *Summary of results*, and *Reporting (the) result*. General background information (overall purpose, methodology and principal results) seemed covered by the first two moves, but no specific research questions or hypotheses were indicated. Her third independent move corresponded to the *Restating results (DCR2)* step in the present study. It is equally noteworthy that her *Reporting (the) results* label overlooked the difference between reporting (stating) and restating (stating again) results, confusing the primary purposes of reporting results in the methods section and reminding the reader of a specific result in the discussion section.

The results suggest that following the two preparatory restatement steps, the majority of the authors of the ERC articles took three key steps to comment on results, *Interpreting results (DCR3), Comparing with previous literature (DCR4)* and *Accounting for results (DCR5).* In this respect, the finding is in line with numerous previous studies (Amnuai & Wannaruk, 2013b; Basturkmen, 2009, 2012; Kanoksilapatham, 2005; Peacock, 2002; Yang & Allison, 2003). One thing that needs clarification is that, besides justifying results, references to previous literature may be made to support explanations too, as Basturkmen (2009) observed.

The results also suggest that in commenting on specific findings, the writers on occasion took additional steps to exemplify their interpretations or explanations (DCR6), or recommend future research (DCR7). Except for Kanoksilapatham (2005, 2015) who identified an exemplifying step in biochemistry and engineering articles, no existing literature mentions these two steps. A subsequent scrutiny of the texts containing the exemplification step confirmed the assumption that this step was unique to qualitative articles. The recommendation of future research step was similar to but differed from that of the *Making deductions from the study (DDS)* move in that the former was indicated by an individual result while the latter was in view of the entire study. In addition, the recommendation step was found not unique to either quantitative or qualitative studies. A plausible reason for its absence in previous literature is that this step did exist but was identified as a constituent step of the

Making deductions from the study (DDS) move. In any event, these two steps seem typical of education research articles.

With respect to the *Evaluating the study (DES)* move, it was observed in two thirds of the discussion sections of the ERC articles. It is interesting that the majority of the authors did not indicate the significance of the study (DES1) while over half of them acknowledged limitations of the study. These results largely support findings of previous studies on the discussion section, which unanimously found the evaluation move in a range of disciplines including applied linguistics, education technology, and biochemistry (Amnuai & Wannaruk, 2013b; Kanoksilapatham, 2005; Yang & Allison, 2003). The results did differ from some previous studies in some ways. Although the move was very common in the ERC articles as in Kanoksilapatham's (2005) biochemistry and Pho's (2008a) applied linguistics texts, it has a very thin distribution in Yang and Allison's (2003) and Amnuai and Wannaruk's (2013b) applied linguistics articles. Furthermore, the indicating significance step was found either absent or rare in Pho (2008a), Kanoksilapatham (2005) and Amnuai (2013b). Nevertheless, despite the differences, it is clear that, regardless of discipline, researchers tend to evaluate their research significantly more by acknowledging limitations than by indicating significance of the study. Perhaps, this can be understood by the mentality of researchers that it is advantageous to critique their own work than have the limitations pointed out by reviewers or readers. An alternative explanation is that while overtly acknowledging limitations to warn the reader

against over-generalization of findings, the authors had a hidden purpose to achieve, i.e., to justify the choices they made in the conduct of the study. Indeed, when identifying limitations, the researchers were actually indicating that the limitations were due to factors beyond their control.

As expected, an overwhelming majority of the ERC article authors employed a *Making deductions from the study (DDS)* move towards the end of the discussion section by either or both of two steps: *Drawing implications or making suggestions (DDS1)* and *Recommending future research (DDS2)*. However, nearly one third of the authors did not recommend further research whereas two thirds drew implications from their study. This finding is in agreement with those of previous researchers who noted a tendency for researchers to refrain from recommending future research (Kanoksilapatham, 2005; Yang & Allison, 2003). As suggested by Berkenkotter and Huckin (1995), in the competitive academic world, some researchers might want to keep potential research topics to themselves.

Due attention must be paid to the last move *Summarizing the study (DSS)*. This move appeared in one third of the ERC articles either at the close of a discussion section or in a separate section of its own under a heading such as "Conclusion" and "Summary." A short summary says goodbye to the reader and offers a take-home message, i.e., the gist of a study. Day and Gastel (2006) advise that to increase chances of a paper being accepted, one should end a discussion with a brief summary.

Moreover, besides a relatively large number of moves and their constituent steps, move recycling contributed greatly to the complexity of the discussion sections investigated. In particular, the four optional and conventional steps of the *Commenting on specific results (DCR)* move, recurred at an astonishing rate of thirteen times per text. In other words, they formed a recurrent cycle of restating hypotheses or research questions – restating results – interpreting results – comparing results with literature – accounting for results. Additionally, as revealed by the high frequency multi-move sequences, this commentary cycle often formed larger cycles with either the *Evaluating the study (DES)* or *Making deductions from the study (DDS)* moves. The cyclic nature of the discussion section has been noted in research articles in many disciplines (Amirian et al., 2008; Basturkmen, 2009, 2012; Lim, 2010; Yang & Allison, 2003). The explanation for move cyclicity in the results section holds true for the discussion section too. In education research, multiple research questions, variables and instruments for data collection typically lead to multiple results to be reported (in the results section) and commented on (in the discussion section). The more questions or hypotheses, the greater cyclicity of rhetorical moves.

4.3 Summary

The first part of this chapter reported results in answer to the first two research questions. The 16 rhetorical moves and their constituent steps identified in the ERC corpus were described in detail, and accordingly, an optimal generic move structure for each section of education research articles was proposed with indications of the most likely order of the moves and steps, and the status of each of them. In the second part, results were interpreted, their underlying reasons contemplated, and tentative claims made, all with reference to previous literature or knowledge of the discourse community of education researchers.

Chapter 5 will present and discuss the results of the MD analysis in answer to Research Questions 3-4.

CHAPTER 5

RESULTS AND DISCUSSION II:

INTER-MOVE LINGUISTIC VARIATION

This chapter reports and discusses the results of the MD analysis in response to Research Questions 3 and 4. Section 5.1 presents the results in two subsections. Subsections 5.1.1 responds to Research Question 3, i.e., What are the co-occurrence patterns of linguistic features in education RAs as represented by the ERC? It presents and interprets the 7 functional dimensions underlying the co-occurring linguistic features identified by a factor analysis. Subsection 5.1.2 answers Research Question 4, i.e., What are the similarities and differences among the moves with respect to the use of the co-occurrence patterns of linguistic features? The similarities and differences are revealed by each move's position on each dimension relative to other moves. Subsequently, Section 5.2 briefly discusses the results.

5.1 Results

5.1.1 Linguistic Dimensions and Their Functions

In answer to Research Question 3, a Principal Component Analysis (PCA) with Promax rotation extracted 7 principal components that account for 36.6% of the total variance of the 36 linguistic features (67 original features) in the ERC corpus.

Table 5.1 Dimensional Structure of Linguistic Features

Dimension 1	Loadings	Dimension 5	Loadings
Mean wordlength	.697	Personal pronouns	.562
Present tense	.587	*To*-infinitives	.499
Nominalizations & gerunds	.515	Present tense	.481
Attributive adjectives	.492	Pro-verb *do*	.389
Phrasal coordination	.344	Relative clauses	.352
Past tense	-.746	Nouns	-.343
Dimension 2	**Loadings**	Participial clauses	-.331
Be as main verb	.845	**Dimension 6**	**Loadings**
Predicative adjectives	.841	Present tense	.566
Pronoun *it*	.489	Existential there	.461
Present tense	.409	Adverbial clauses	.425
Pragmatic expressions	.324	Present perfect aspect	.397
Complements	.307	Negation	.389
Past tense	-.423	Adverbs	.315
Dimension 3	**Loadings**	Past tense	-.395
Dispreferred forms	.620	**Dimension 7**	**Loadings**
Modals	.509	Demonstratives	.488
Passives	.483	Prepositional phrases	.451
Adverbs	.401	Independent clause coordination	-.559
Pronoun *it*	.386	Past tense	-.460
To-infinitives	.357	Type/token ratio	-.422
Nouns	-.619		
Prepositional phrases	-.412		
Dimension 4	**Loadings**		
Complements	.749		
Public verbs	.628		
Suasive verbs	.539		
Private verbs	.412		
Pragmatic expressions	.318		

Table 5.1 above, based on the structure matrix, summarizes the factorial

structure of the linguistic features. Each of the seven dimensions has at least five

features loaded on it, with absolute values ranging from .307 to .845. With the exception of Dimension 4, all dimensions have both positive and negative loadings.

The loading of a given feature quantifies the correlation between that feature and the factor on which it loads. In other words, it indicates the strength of the co-occurrence relationship between the feature and the rest of the features as a whole on the same factor. The absolute value of a loading determines its magnitude. The higher the value, the stronger the correlation. For instance, -.50 indicates a stronger correlation than .45. Positive and negative values point to a likely complementary pattern of the distribution of groups of features. That is, the presence of co-occurring positive features predicts the tendency for the negative features to be absent in the text, and vice versa.

The key to interpreting the basic function of a factor lies in understanding the communicative function most widely shared by the set of co-occurring linguistic features defining a factor as a dimension of variation (Conrad & Biber, 2001). It is crucial to identify what motivates the co-occurrence patterns of a set of features, positive or negative. Similarly, in the presence of negative features, it is also necessary to know the reasons for the complementary distribution of the positive group and the negative group.

The following sections present and interpret the dimensions in terms of their basic, rather abstract communicative functions without making direct reference to

their realizations in the rhetorical moves in the ERC articles. A label is proposed to characterize each of them in accordance with the interpretations. Admittedly, the interpretation of the negative ends of Dimensions 1, 2, and 6 is inevitably tentative because these ends are marked by only a single feature.

5.1.1.1 Dimension 1: Current Information vs. Procedural Concerns

Dimension 1 has five positive and one negative loadings in complementary distribution, explaining the biggest portion of the total variance (9.3%). The positive loadings are mean word length, present tense, nominalizations & gerunds, attributive adjectives, and phrasal coordination. The negative feature is past tense. Present tense and past tense cross-load on a few other dimensions.

A focus on current information is clearly discernible of the features on the positive end. Word length, which loads the highest on the dimension, marks great density of information because longer words have more specific, specialized meanings than shorter ones. Nominalizations & gerunds are also informative. Being longer nouns derived from verbs, they are characterized by referential specification (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Chafe (1982, 1985) points out that nominalizations condense information by expanding idea units and presenting information in fewer words. Attributive adjectives, either pre- or post-modifiers of nouns, further elaborate nominal information, contributing to information density. Phrasal coordination also plays an important role in integrating information as phrases

are linked by means of coordinating conjunctions such as *and*, *or*, and *but*. The other positive feature, present tense, loads the second highest on the dimension. According to Biber (as cited in 1988), present tense verbs are generally associated with topics and actions of immediate relevance, and in academic discourse, they have a focus on information being presented without indicating temporal sequencing. Together, the five co-occurring features exhibit a marked an informational focus in text.

The negative end of Dimension 1 has past tense as the only feature on it. Its interpretation is therefore not as easy. The principal use of past tense verbs is to convey an action or situation in past time. As such, they are the most important marker of narrative text (Biber, 1988). Importantly, unlike present tense verbs, they may indicate temporal sequencing. This property may highlight a concern over the research procedures in research articles.

In sum, Dimension 1 functions as a parameter of textural variation that marks where a text is on a continuum from a focus on current information to a focus on narrating past actions or states, hence the label *Current information vs. procedural concerns.*

5.1.1.2 Dimension 2: Evaluative Stance vs. Past Actions or States

Dimension 2, which contributes 6.7% to the total variance, comprises six features with positive weights and one with a negative weight. *Be* as main verb, predicative adjectives, pronoun *it*, present tense, pragmatic expressions and

that-complements load on the positive end while past tense is located on the negative end. Among them, present tense and past tense, with moderate weights, cross-load on a few other dimensions.

The shared function of communicating evaluative stance is responsible for the high co-occurrence rates of the six features positively loaded on Dimension 2. *Be* as main verb and predicative adjectives are two greatest loadings on the dimension. A form of the main verb *be* is frequently followed by a predicative adjective. Predicative adjectives frequently function as heads of *that* or infinitive complements to convey stance (Winter, as cited in Biber, 1988). Complements, including *that*-verb complements, *that*-adjective complements, Wh-clauses, and infinitive complements, are known as expressive of evaluative stance. Winter (as cited in Biber, 1988) notes that both verb- and adjective-complements facilitate the expression of value assessment of the propositional information in the dependent clause, with the evaluation placed in the main clause and the information stated in the *that*-clause. While *that*-complements fulfill the function of evaluating statements, What-clauses (restricted to those complementing verbs in the present study), express evaluative stance towards an indirect question in the form of a Wh-clause (Winter, as cited in Biber, 1988). Pragmatic expressions, which include downtoners, hedges, amplifiers and emphatics in the present study, have been noted to frequent scientific discourse expressing the writer's evaluative or commitment stance towards propositions.

Downtoners mitigate the force of a verb (e.g., *almost*, *somewhat*, *practically*), whereas hedges (modals not included, e.g., maybe, *about*, *perhaps*) withhold the writer's full commitment to a proposition by indicating uncertainty and tentativeness. In contrast, amplifiers (e.g., *absolutely*, *altogether*, *thoroughly*) boost the force of a verb while emphatics (e.g., *just*, *for sure*, *in fact*, *such a* + adjective) indicate the extent of certainty. Hyland (2004), who groups downtoners and hedges as hedges, and amplifiers and emphatics as boosters, considers hedges and boosters evaluative devices that reflect the writer's assessment of his or her own or others' propositions. As for the moderately weighted feature of present tense, it is quite obvious that present tense verbs are used in association with the above features, for example, as heads of the complements or in the propositions hedged or boosted.

The negatively weighted feature, past tense, forms the other end of the continuum of Dimension 2. As on Dimension 1, past tense verbs seem to mark an orientation towards past actions or states.

Taken together, Dimension 2 spans from expressing evaluative stance to narrating past actions, hence the label *Evaluative stance vs. past actions or states*.

5.1.1.3 Dimension 3: Logical Probability vs. Integrated Information

Six features cluster on the positive and two on the negative end of this dimension, accounting for 5.1% of the total variance. The two sets form a complementary relationship.

The positive end of the dimension is largely defined by modals and passives. The highest loading is dispreferred forms followed by modals and passives. Dispreferred forms themselves, in fact, largely involve modals and passives. In the present study, the dispreferrred forms feature is five original features combined because most of them are extremely rare in academic writing and because they come under the same category of reduced forms and constructions dispreferred by prescriptive grammarians (Biber, 1988). In the ERC corpus, contractions, subordinate *that* deletion, stranded prepositions, and split infinitives are almost completely absent, leaving split auxiliaries virtually representing the whole category. A split auxiliary is defined as an auxiliary (mostly modals) followed by an adverb which is followed by a verb (e.g., *could perhaps improve, may then be needed*). Therefore, the functional focus of this end is mostly represented by modals and passives. Biber et. al (1999) state that modals mostly have stance-type meanings and that *could, may*, and *might* almost exclusively mark logical probability in academic writing with varying degrees of uncertainty or tentativeness. They also note the frequent patterning of passive voice and modals in academic prose. As to the function of passive voice, Biber and his colleagues state that in academic writing, passive voice has an important function of demoting the agent thus giving topic status to the patient. It is frequently used to report findings and express logical relations as well as to describe aspects of scientific methodology and data analysis procedures. In addition, many adverbs such as

generally, *possibly*, *probably*, *certainly*, and *obviously* are linked with the author's attitude towards the proposition in the clause, its generalizability or its expectedness. In academic writing, besides substituting for nouns and phrases, pronoun *it* often functions as anticipatory subject/object or subject in cleft constructions to refer to propositions expressed by clauses. It is natural for pronoun *it* to co-occur with the other features in the set since the multiple uses of pronoun *it* can efficiently help the writer express logical possibilities of a prediction, result, or knowledge claim. The last feature on the positive end, *to*-infinitives, also contributes to the expression of logical probability by integrating information and expanding idea units which complement nouns, adjectives, and verbs, (Chafe, 1982, 1985).

On the negative end, nouns and prepositional phrases co-occur to mark an information focus. While a high frequency of nouns (nominalizations & gerunds are counted separately) marks conceptual abstractness, prepositional phrases pack large amounts of information expressed by nouns (Chafe, 1982, 1985). Together, they make the discourse highly informational.

The label *Logical probability vs. integrated information* is suggested for Dimension 3 as it measures the extent to which a text communicates logical probability or has an informational focus.

5.1.1.4 Dimension 4: Commentary

This dimension explains 4.1% of the total linguistic variation. Unlike the other dimensions, it has only positive features loaded on it, hence a monopolar dimension.

An orientation towards commentary readily emerges from the co-occurrence of complements, verbs of different semantic categories with specific communicative functions, and pragmatic expressions. Complements, particularly *that*-complements following a verb, express propositions that are commented on. The three different classes of verbs comment on the propositions, reporting, explaining, and interpreting them. Public verbs, those expressing observable actions, are primarily speech act verbs (e.g., *claim*, *report*, *say*) that state the propositions in reported speech; private verbs, aka verbs of cognition (e.g., *believe*, *assume*, *think*), express intellectual states or unobservable intellectual acts (e.g., *discover*, *conclude*, *demonstrate*); suasive verbs (e.g., *recommend*, *suggest*, *propose*) intend to persuade, recommend, or urge the reader to accept what is expressed by the *that*-complement (Biber, 1988). To achieve effectiveness, the writer resorts to pragmatic expressions such as downtoners, hedges, amplifiers and emphatics to mitigate the force of assertion, or to indicate uncertainty and tentativeness thus sounding polite and moderate, or to emphasize an argument.

This dimension shares with Dimension 2 two features, i.e., complements and pragmatic expressions. However, they are markedly different in focus. The former is mostly defined by complements and verbs which together express explanations and interpretations; the latter, the positive end of which is characterized by different forms of *be* followed by predicative adjectives, basically evaluates what is expressed in complements.

In sum, this dimension expresses opinions about a systematic series of explanations and interpretations of propositions. It is therefore named *Commentary*.

5.1.1.5 Dimension 5: Personal Engagement vs. Modified Information

Dimension 5 has personal pronouns, *to*-infinitives, present tense, pro-verb *do*, and relative clauses loaded on its positive end; and nouns and participial clauses on its negative end. It contributes 4.0% to the total variance.

At first sight, the highest loading of personal pronouns appears to mark an interpersonal focus. However, a preliminary analysis of the ERC corpus suggested the presence of third person plural pronouns (*they, them, themselves, their*) and first person pronouns (*I, me, my, myself, we, us, ourselves, our*) and the nearly complete absence of third person singular pronouns (*he, him, his, himself, she, her, herself*) and second person pronouns (*you, your, yourself, yourselves*). That is why in the first place they were combined into one single feature. Therefore, third person plural and first person pronouns are considered to largely define this pole of dimension. In

written texts, while third person plural pronouns refer to a specific group of individuals identified in the preceding text, first person pronouns substitute for the writer or writers. In the research article genre, third person plural pronouns can refer to a wide range of entities, for example, questions and previous studies, but mostly have human referents, whether participants or the population that the participants represent. *To*-infinitives, which frequently co-occur with personal pronouns, indicate intentions, purposes, desires, efforts, perceptual states, and various types of other general actions (Biber, Johansson, Leech, Conrad, Finegan, et al., 1999). Pro-verb *do* also expresses actions conveyed by a specific or a series of lexical verbs, thus marking personal involvement in an action. In general, present tense, as described previously, deals with actions of immediate relevance, expressing intentions, purposes, desires, efforts and other types of actions (Biber, 1988). The other feature, relative clauses, including *that*-, Wh-, pied-piping, and sentence relatives, are used for more exact and explicit reference in writing (Biber, 1988), serving as a device for integration and idea unit expansion (Chafe, 1982, 1985). Overall, when they co-occur, these features mark a focus of personal engagement, conveying actions to be undertaken by the author or groups of other people.

The two features, nouns and participial clauses, on the negative end have rather low weights. They have a focus on integrated information. As previously described, high frequencies of nouns are indicative of large amounts of information. It

is more the case when followed by participial clauses. Participial clauses, significantly more common in writing, modify the information expressed by the nouns, giving the information a more limited or restricted quality.

In summary, the positive end of Dimension 5 measures the extent to which texts vary in terms of actions involving people while the negative end the extent to which texts exhibit an informational focus. Thus, the label *Personal engagement vs. modified information* is proposed for this dimension.

5.1.1.6 Dimension 6: Unsatisfactory Status Quo vs. Research Conduct

Responsible for 3.7% of the total linguistic variation, the set of present tense, existential *there*, adverbial clauses, present perfect aspect, negation and adverbs on the positive end, and past tense alone on the negative end, form a complementary relationship.

Clearly, the underlying construct of the positive end is associated with the existing state of affairs being challenged for change. Both present tense and present perfect aspect are relevant to the present state of affairs. While the former designates an action or state of affairs of the present time, the latter describes an event or state that occurs during a period of time in the past and extends up to the present time. Moreover, Biber et al. (1999) note the tendency in the scientific genre to report the views or findings of authors of previous research using the simple present tense. They also state that in academic writing, the present perfect aspect typically implies a continuing validity of earlier findings or practices. Existential sentences, with *there* as

a dummy subject and *be* as the main verb placed before the real subject, assert the existence or presence of something. When used with the simple present tense or present perfect aspect, existential sentences refer respectively to the existence or presence of something at the present and something that began to exist in the past and continues to exist until the present time. Overall, adverbial clauses indicate informational relations conveying time, condition, reason, cause, manner, comparison, or concession (Quirk, Greenbaum, Leech, & Svartvik, 1985). Adverbs, as modifiers in adjective and adverb phrases or as adverbials in the clause, most typically denote degree, or the circumstances under which an action, process, or state takes place, or the connection between clauses, or the writer's attitude towards information (Biber, Johansson, Leech, Conrad, Finegan, et al., 1999). The writer's dissatisfaction with the current state of affairs emerges when the above features co-occur with negative words (*not*, *no*, *neither*, *nor*) that deny or reject something under discussion.

On the negative end of the dimension is past tense, the most important marker of narrative text (Biber, 1988). As already mentioned, past tense verbs convey an action or situation in the past, often indicating temporal sequencing. It can be interpreted here as narrating the execution of research aimed at bringing about changes in the status quo of affairs.

The label *Unsatisfactory status quo vs. research conduct* is thus proposed to capture the functional characteristics of Dimension 6.

5.1.1.7 Dimension 7: References to Present Research vs. Information

about the Past

Dimension 7 is represented by demonstratives and prepositional phrases

on the positive end, and independent clause coordination, past tense and type/token

ratio on the negative end. It explains a total of 3.6% of the total variance.

The positive end is marked by frequent references to the research itself.

Demonstratives (*this*, *that*, *these*, *those*) can refer to entities outside the text, and

particularly in writing, to specific nominal entities or to inexplicit, often abstract,

concepts in the text. Coupled with prepositional phrases which pack large amounts of

information expressed by nouns (Chafe, 1982, 1985), demonstratives make it possible

for the writer to conveniently discuss different aspects of the present study mentioned

near or far away in the text, achieving what Halliday and Hasan (1976) termed as

referential cohesion. Entities can include the research itself as a whole or aspects of

the research such as research questions, participants, results, etc.

The negative end seems to suggest integrated information relevant to past

time. Independent clause coordination is restricted to *and* as a general purpose clause

coordinator capable of marking many different logical relations between two clauses

(Biber, 1988), but basically it adds a new idea unit to a preceding one. A high

type/token ratio is interpreted as marking a highly exact presentation of information,

conveying maximum content in the fewest words. The higher the ratio, the greater

variety of word types and the higher the amount of information integrated. Together with past tense, these features characterize the negative pole of the dimension as expressing condensed information with respect to past time.

This dimension is therefore named *References to present research vs. information about the past*.

5.1.2 Inter-move Linguistic Variation

In answer to Research Question 4, this subsection presents the similarities and differences among the moves on the seven functional dimensions underlying the co-occurrence patterns of linguistic features. Table 5.2 below summarizes the results of the seven one-way ANOVAs aimed at comparing the mean dimensions scores of the moves with respect to the seven dimensions.

The results show that at the $p \leq 0.5$ significant level, significant differences exist among the moves with respect to all dimensions, F = 13, $p = 0.01$, $r_1 = 0.63$, $r_2 = 0.45$, $r_3 = 0.32$, $r_4 = 0.32$, $r_5 = 0.32$, $r_6 = 0.45$, $r_7 = 0.45$. This indicates that overall, the 14 moves are linguistically distinguishable on all dimensions. That is to say, each of the dimensions is a reliable parameter of linguistic variation among the moves.

Table 5.2 ANOVA Results of Move Differences along the Dimensions

Dimension	*r*	df	F	*p*
1	0.63	13	143.0	0.01
2	0.45	13	51.9	0.01
3	0.32	13	29.4	0.01
4	0.32	13	28.5	0.01
5	0.32	13	31.2	0.01
6	0.45	13	76.9	0.01
7	0.45	13	69.8	0.01

Note. r = effect size; df = degree of freedom; F = F-ratio; p = p-value

Due to violations of homogeneity of variance and the dramatic differences among the moves in sample size, no *post hoc* analyses were performed to establish specifically which pairs of moves are significantly different and which ones are similar. In the following subsections, similarities and differences of the moves are instead discerned by comparing their mean dimension scores on each of the seven dimensions. Figures 5.1-5.7 visualize relations among the moves. On each dimension, moves some distance away from other moves are different whereas those clustering together are similar.

As a rule, three examples with their scores on the dimensions are provided to illustrate the characteristics of each dimension: one representing the move at the positive extreme, one the move on the negative extreme , and one the move closest to the mean (marked "0" on the dimensions). For better illustrative effects, move texts with a higher dimension score than the mean score may be used in some cases to illustrate moves at the extremes of the dimensions.

5.1.2.1 Variation along Dimension 1

Figure 5.1 presents the mean dimension scores of the moves on Dimension 1 labeled *Current information vs. procedural concerns*. Four clusters of moves are readily visible on this dimension, each remotely separated from other clusters or individual moves, indicating a shared communicative function within each group, and at the same time, broad variations among the groups.On the positive end of the dimension, the move with the highest positive mean*, Establishing a niche (IEN),* forms a cluster with two other moves, *Summarizing the study (DSS)* and *Establishing a territory (IET).* United by their close means, the three moves are marked by great density of current information realized by frequent co-occurrences of high values of word length, present tense, nominalizations &
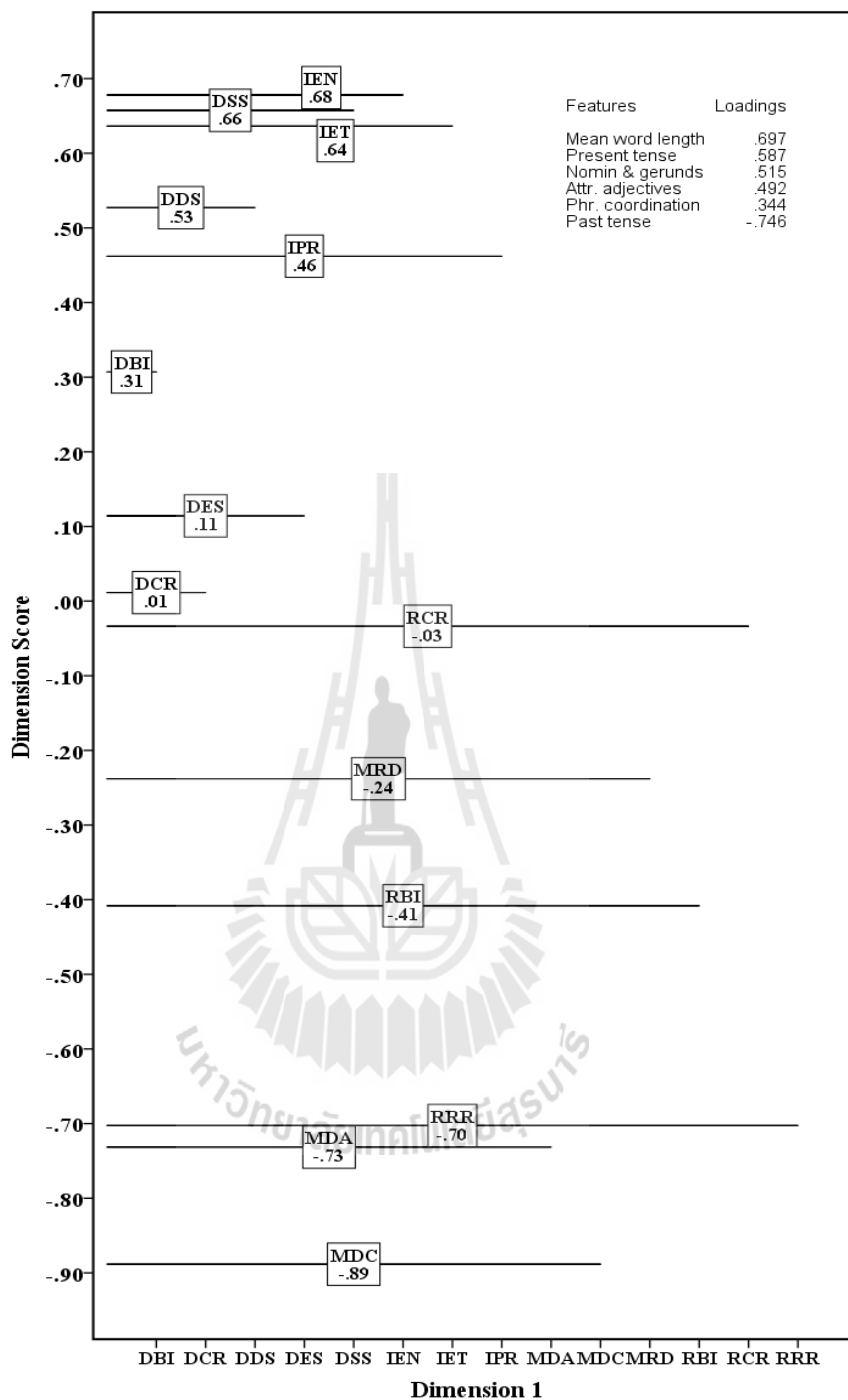
Figure 5.1 Moves along Dimension 1

gerunds, attributive adjectives, and phrasal coordination. At the same time, past tense

verbs are markedly rare in these moves, indicating the absence of procedural concerns.

The following text sample, an *Establishing a niche (IEN)* move, illustrates these

characteristics, with present tense enclosed in (…), nominalizations & gerunds in […],

attributive adjectives in {…}, and phrasal coordination in <…>.

>   *1) The [arguments] presented above (provide) contradicting [possibilities]. The [irrelevance] criticism (suggests) there (is) likely no [relationship] between CEO MBA [education] <and> CEP. The profits-first criticism of MBA [education] (suggests) a {negative} [relationship] from the {normative} perspective of CEP, <but> a {positive} [relationship] from the [business]-case perspective. Depending on which [argument] (is) supported, drastically {different} [implications] will ensue. If no [relationship] (is) found, the [irrelevance] criticism would be the likely [explanation]. If a {negative} [relationship] (is) discovered, the profits-first criticism would be supported, implying that MBA [education] (erodes) students' sense of {moral} <and> {social} [responsibility]. Finally, if the [relationship] (is) positive, MBA [education] could be said to actually have a {positive} outcome for all stakeholders involved (i.e., {triple} bottom-line) <and> would at the very least refute some of the [irrelevance]criticisms. ....*
>   *.. (LAE_09) (D1 Score = 2.61)*

This text has at least three observations of each of the co-occurring

features on the positive end of Dimension 1, indicating a large amount of information

involved. Simultaneously, past tense is completely absent from the text, suggesting

that the text is irrelevant to procedural concerns. The author establishes a niche for his

or her research by analyzing the conflicting arguments over the value of MBA

education. Outcome predictions of an proposed investigation are made based on a

large amount of information indicated by the co-occurring features. These linguistic

devices serve the author's purpose very well.

>   In contrast, the *Describing data collection procedures (MDC)* move sits

alone at the farthest end of the negative pole, suggesting that it is distinctly marked by

procedural concerns with high frequencies of past tense verbs, as illustrated by the

following excerpt.

> *2) The examiner first (modeled) the task using English letters to ensure that the directions (were) understood. Children (were) taught to use letter blocks to spell the initial and final sounds of the following words: fur, pear, jar, and rip. The five sounds modeled (were) not target sounds. Then the test words (were) pronounced and children (used) Hebrew letter blocks to spell their initial and final sounds. The number of words spelled correctly in each set (**Shehzad**) scored...............* .... *(SSR_02) (D1 Score = -1.19)*

In the excerpt, all clauses contained a past tense verb (enclosed in (…)) to describe data collection procedures in chronological order. Meanwhile, the text segment does not seem to have a focus on current information as it relatively lacks those co-occurring features marking the positive end. There are four instances of attributive adjectives (*initial* and *final* repeated twice) but there is only one nominalization (*directions*), and there is no phrasal coordinator or present tense verb. Besides, as can be seen, this segment contains many short words.

Another three moves, including *Commenting on specific results (DCR)* and *Evaluating the study (DES)* of the discussion section and *Commenting on specific results (RCR)* of the results section, form a group around the dimension mean. Neither do they have a clear-cut focus on current information, nor do they function to narrate procedures. The following is an excerpt from a *Commenting on specific results (DCR)* move with positive features in (…) and negative features in […].

> *3) Unfortunately, the student, perhaps relying simply on the key word left to select the (subtraction) (operation), then [failed] to represent the structure of the second-step change problem correctly, which (requires) (addition) because the missing (information) (occurs) in the first position of the change (equation) with a decreasing change value...(ESJ_05)(D1 Score = 0)*

Apparently, the text is mixed in that neither positive nor negative features are predominant. There are two positive features present in the text: five nominalizations and two present tense verbs; the other positive features, namely, attributive adjectives and phrasal coordinators are totally absent. In addition, its average word length appears to be moderate. Additionally, the presence of the negative feature, past tense, further neutralizes its focus on current information. In consequence, this text focuses neither on current information nor on past procedures. This analysis of linguistic features on Dimension 1 provides further evidence that this text's communicative purpose is to comment on a specific result, in this case, via accounting for the result.

The fourth group consists of *Making deductions from the study (DDS)* and *Presenting the present research (IPR)* situated some distance from the first group on the positive end. Their moderately high mean dimension scores point to a moderate degree of informational density.

Also on the negative pole, with little difference in means, *Describing data analysis procedures (MDA)* and *Reporting specific results (RRR)* form a pair marked by a relatively high degree of procedural concerns.

The other three moves, viz., *Providing background information (DBI)* on the positive end and *Describing research design (MRD)* and *Providing background information* (RBI) on the negative end, do not group with any other move. This

indicates that they each differ significantly from other moves. While DBI, with a

moderate positive mean, is moderately marked by dense information, MRD and RBI

are moderately procedure-relevant due to their moderate negative means.

5.1.2.2 Variation along Dimension 2

The relations among the moves with respect to Dimension 2, labeled

Evaluative stance vs. past action, are summarized in Figure 5.2 below.

Clearly, *Commenting on specific results (RCR)* and *Evaluating the study*

(DES) form a pair with the highest means at the positive extreme. They are

significantly different from any other group or individual move. The following is a

sample of an RCR move, illustrating how the co-occurring features together

communicate the author's stance towards propositional content. These features

include *Be* as main verb (in (…)), predicative adjectives (in […]), pronoun *it* (in {..}),

present tense (in <…>), and *that* complements (underlined).

> *4) {It} (is) also [possible] <u>that</u> our students' lack of work experience*
> *could have the effect of channeling their efforts in directions that (were) [prone] to be*
> *more [individualistic] in nature as opposed to more [collectivistic] in nature.*
> *...(LAE_05)(D2 Score = 6.91)*

This one-sentence excerpt exhibits virtually all features indicative of the

underlying construct, that is, evaluative stance, which marks the positive pole of the

dimension. The sentence is structured with pronoun *it* as the anticipatory subject
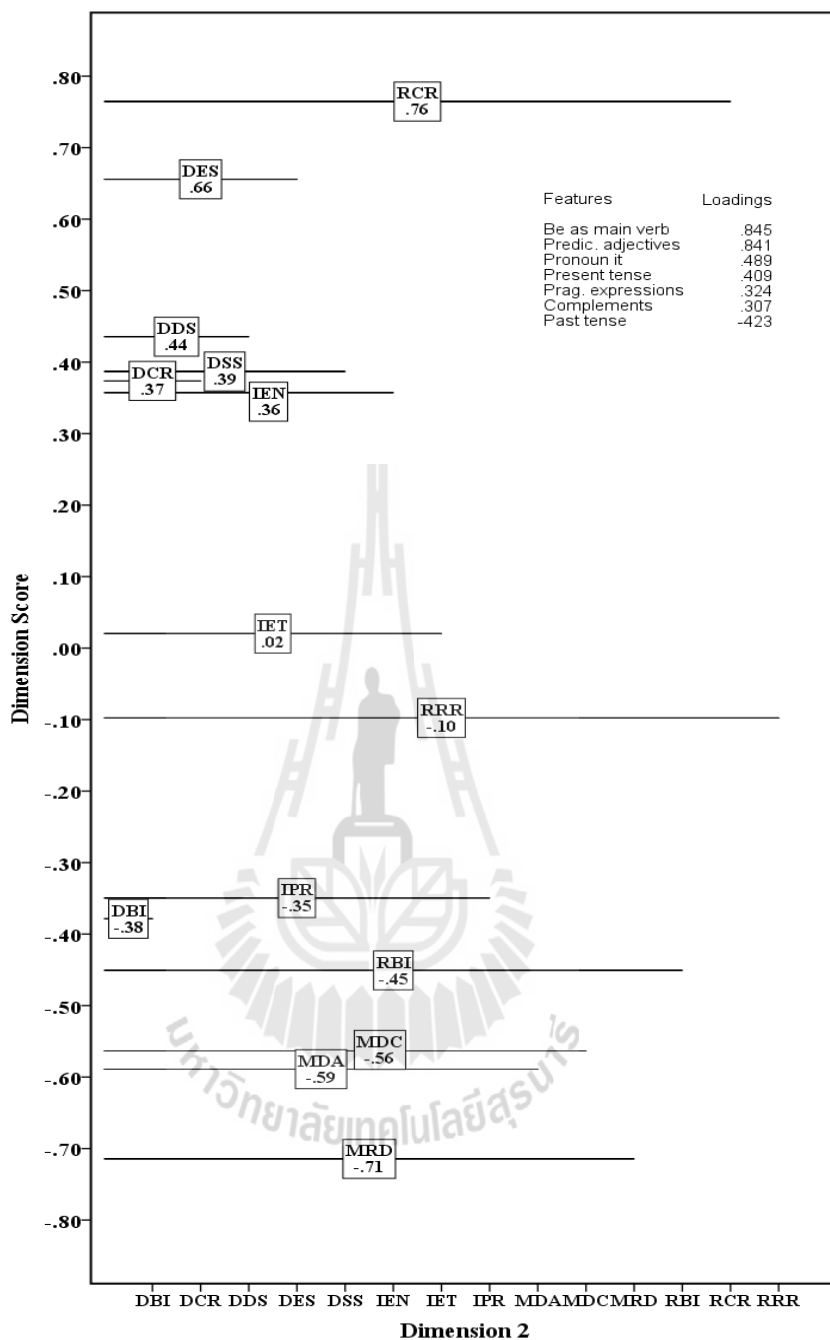
Figure 5.2 Moves along Dimension 2

followed by *is* as the main verb and the adjective *possible* as the predicate which in

turn heads a *that*-complement. The sentence conveys the author's evaluative stance

towards the cause of students' lack of work experience, with the "possible" cause

identified. In addition, within the *that*-complement, three more predicative adjectives and *were* as the main verb further contribute to the evaluative stance. As revealed in 4.1.4.3, the communicative purpose of the *Commenting on specific results (RCR)* move in the results section entails the author's explanation for a particular result just presented and expressing personal stance is inevitable.

At the farthest negative end is *Describing research design (MRD)* of the methods section, marked by a focus on past actions or states expressed with past tense verbs. As illustrated in the following text of an MRD move, past tense plays a vital role in realizing this underlying basic communicative function. In this case, the author uses six past tense verbs within four sentences to clarify what things were like and what was done in relation to the two vital aspects of the research design—type of research design and variables involved.

5) This study [employed] a quasi-experimental design for cross-sectional data in which the high school mathematics curriculum a student completed [was] the independent variable of most interest. Difficulty of students' first university-level mathematics course and the grade earned in that course [served] as dependent variables. This quasi-experimental design [allowed] effects corresponding to the research question to be examined but [did] not automatically permit strong causal inferences because of the absence of random assignment. As a result, our goal [was] to statistically control for important differences between the two curriculum groups that [could] bias our inferences. ....(AER_02)(D2 Score = -.74)

The *Establishing a territory (IET)* move has a mean score very close to the dimension mean. As the following IET text illustrates, it is oriented neither towards expressing evaluative stance nor towards narrating past actions or states.

6) Students' questions [may] be triggered by inconsistencies between their prior knowledge and the new information which they [are] receiving, which then [engenders] cognitive dissonance (Citation) or "epistemic curiosity" (Citation). For students, posing their own questions {is} a first step towards filling their knowledge gaps and resolving puzzlement. Asking questions [allows] them <to articulate> their current understanding of a topic, <to make> connections with other ideas, and also <to become> aware of what they [do] or [do] not know. .... (IET_JEE_07)(D2 Score =.02)

In the above text, although there are some instances of positive features, including six present tense verbs (in [...]), one *be* as main verb (in {...}), and three infinitive complements (in <...>), no predicative adjectives or pragmatic expressions are found to co-occur with them. Therefore, hardly any evaluative stance is expressed. Likewise, the absence of the negative feature of past tense indicates that the text does not narrate past actions or states. In consequence, the text is not marked on this dimension.

Turning to the most conspicuous bundle of four moves on the positive pole, as seen in Figure 5.2, three of them, viz., *Making deductions from the study (DDS)*, *Commenting on specfic results (DCR)*, and *Summarizing the study (DSS)* are discussion section moves while one, *Establishing a niche (IEN)*, belongs to the introduction section. Being so closely packed, these moves are very similar in that they express a moderate degree of evaluative stance.

On the negative pole, two moves, namely, *Describing data collection procedures (MDC)* and *Describing data analysis procedures (MDA)*, form a pair with very close means. This suggests that they are similarly characteristic of a focus on past actions or states defined by high frequencies of past tense verbs.

Still another group of moves is distinctly visible on the negative pole. Although each belongs to a different section, the three moves, namely, *Presenting the present research (IPR)* of the introduction section, *Providing background information (DBI)* of the discussion section and *Providing background information (RBI)* of the results section, share the basic function of communicating past actions or states. However, having much lower mean scores, they are considerably less marked than those moves with much higher means, for instance, the *Describing research design (MRD)* move at the extreme end of the negative pole.

The remaining *Reporting specific results (RRR)* move is almost unmarked due to its extremely low mean score on the dimension.

5.1.2.3 Variation along Dimension 3

Figure 5.3 presents the distribution of the moves along Dimension 3: *Logical probability vs. integrated information.*

At the positive extreme is *Making deductions from the study (DDS)* of the discussion section with the largest mean. It does not group with other moves and therefore it differs significantly from them in the extent to which it is marked by an orientation towards expressing logical probability. The following DDS move text typifies how the co-occurring features jointly realize this basic communicative function.

7) *<To provide> accessible opportunities for more students <to experience> doing SEM in authentic and exciting communities of practice, regularized partnerships (could) [be developed] <to link> schools to places of SEM*

practice like community colleges and universities, hospitals, museums, and technical laboratories. Such programs (could) allow students <to explore> identity connections to science in their own individual ways. Research on how such informal, active learning settings help students discover their abilities and options and {perhaps} develop a sense of themselves as people who enjoy learning/doing science and engineering (would) contribute to a growing and important body of literature <to guide> policy and practice in this area (Citation)................ (RST_04)(D3 Score = 1.59)
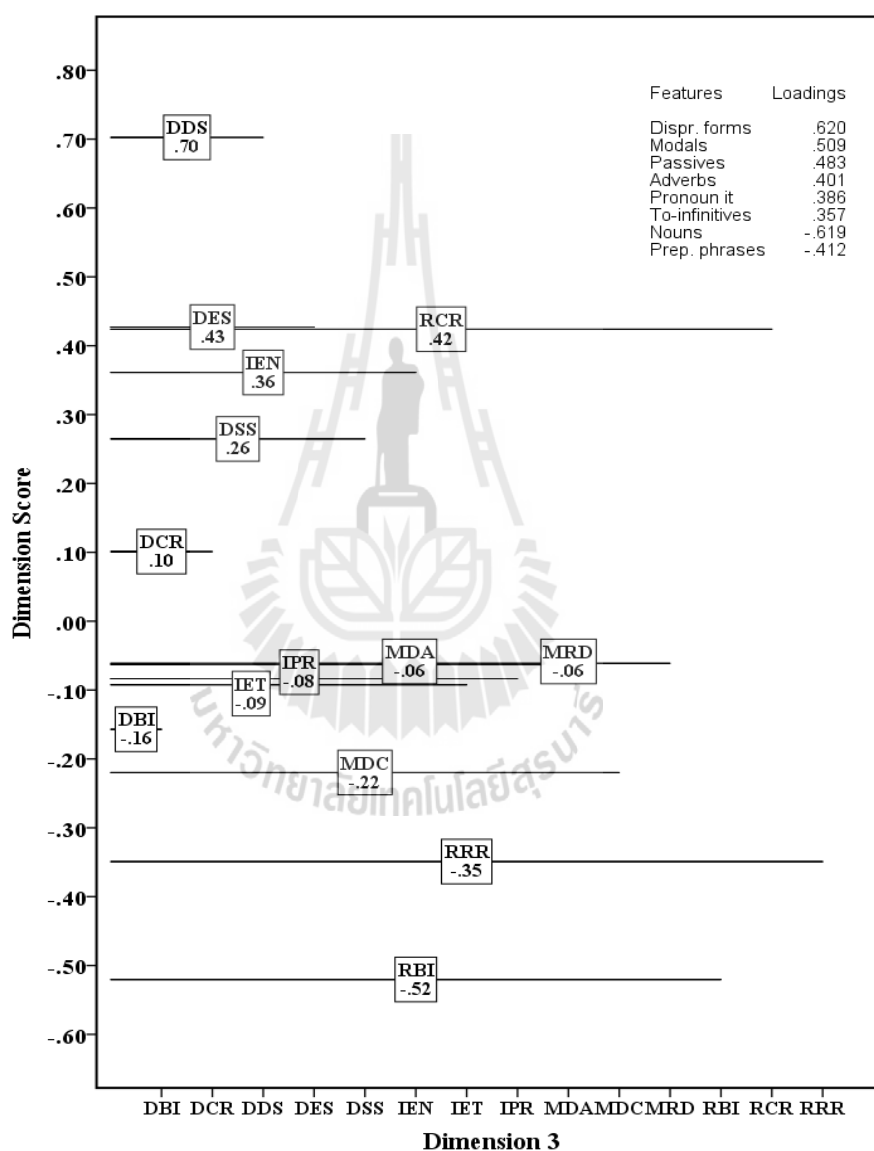

Figure 5.3 Moves along Dimension 3

As discussed previously, the basic function of communicating logical probability is largely realized by modals, passives, *to*-infinitives, adverbs and pronoun

it. In the above text, within only two sentences, there are three modals (in (…)), one passive (in [..]), one adverb (in {…}) and five *to*-infinitives (in <…>). These co-occurring features convey logical relationships by means of linking activities to their possible consequences. The basic, abstract communicative function of the positive pole is clearly embodied in the actual, concrete communicative purpose of this DDS move. In drawing pedagogical implications and recommending future research on the basis of the study, the author suggests what activities be done and what the likely results will be of the suggested activities.

As can be seen, *Providing background information (RBI)* of the results section occupies the farthest negative end of the dimension representing a focus on information realized by the co-occurrence of nouns and prepositional phrases. The following RBI move illustrates how nouns and prepositional phrases together convey this basic function.

8) *To determine the effects of agent race and gender, a series of 2 (Agent race: Black vs. White) X 2 (Agent gender: Male vs. Female) between-groups analyses of variance (ANOVAs) were performed on each of the key dependent variables: gender stereotypes, engineering stereotypes, utility, self-efficacy, and interest. In addition, to determine how the participants felt about the agent and how persuasive the message was, 2 (Agent race: Black vs. White) X 2 (Agent gender: Male vs. Female) between-groups ANOVAs were performed for agent-liking and for message-persuasiveness (see Table 1 and 2 for all analyses). We had predicted that agents that were more similar to the participants in race and gender would be the most effective in changing the participants' beliefs about engineering….. (JEE_01)(D3 Score = -.85)*

The above text abounds in nouns (e.g., *effects* and *utility*) and prepositional phrases (e.g., *on each of the key independent variables* and *in addition*)

and, as anticipated, has a very limited presence of the features marking the positive pole of the dimension. While a large number of nouns mark conceptual abstractness, prepositional phrases consolidate large amounts of information expressed by the nouns into larger information units. This basic function fulfils the communicative purpose of the RBI move, that is, to provide whatever information necessary (in this case, the data analysis procedures and hypothesis) to ensure correct understanding of the results to be presented.

On the positive end, three moves, viz., *Evaluating the study (DES)*, *Commenting on specific results (RCR)*, and *Establishing a niche (IEN)*, form a conspicuous cluster far away from other moves. This suggests that while these moves are similar in the degree in which they are oriented towards conveying logical probability, they are significantly different from the other moves in this respect. Compared with the high mean score of DDS, these moves have moderately low means, indicating that they are considerably less marked by an orientation towards logical probability. In other words, these moves have a less pronounced presence of the defining features, i.e., modals and passives in particular.

Another prominent group of four moves is situated slightly below the dimension mean, suggesting that they hardly exhibit an orientation towards either expressing logical probability or an informational focus. These moves are *Establishing a territory (IET)* and *Presenting the present research (IPR)* of the

introduction section, and *Describing research design (MRD)* and *Describing data analysis procedures (MDA)* of the methods section. The MDA text below typifies these moves on this dimension.

> 9) *Studies assessing {the reliability and validity of the IPAQ} [have been conducted] in 14 centers in 12 countries on six continents using standardized methods. The results demonstrate reasonable reliability and validity, with broad applicability to {a wide range of countries and cultures}. In this study, physical activity [was assessed] at {baseline, midpoint and end of the 12-week intervention period}*...
> .... *(HER_03)(D3 Score = .05)*

The text has two instances of passive voice (in […]) but no other co-occurring features to mark a logical probability orientation. Similarly, there are only three nouns modified by a prepositional phrase (as in {…}) that mark an informational focus. As a results, the text is almost unmarked on this dimension.

The other moves do not group together with other moves due to their varying means. Thus, they are different from the other moves in the degree they are marked by either logical probability or an informational focus. These moves are *Summarizing the study (DSS)*, *Commenting on specific results (DCR)*, *Reporting specific results (RRR)*, *Describing data collection procedures (MDC)*, and *Providing background knowledge (DBI).*

5.1.2.4 Variation along Dimension 4

Figure 5.4 below visualizes the similarities and differences among the moves with respect to Dimension 4 which, being monopolar, is labeled *Commentary.*
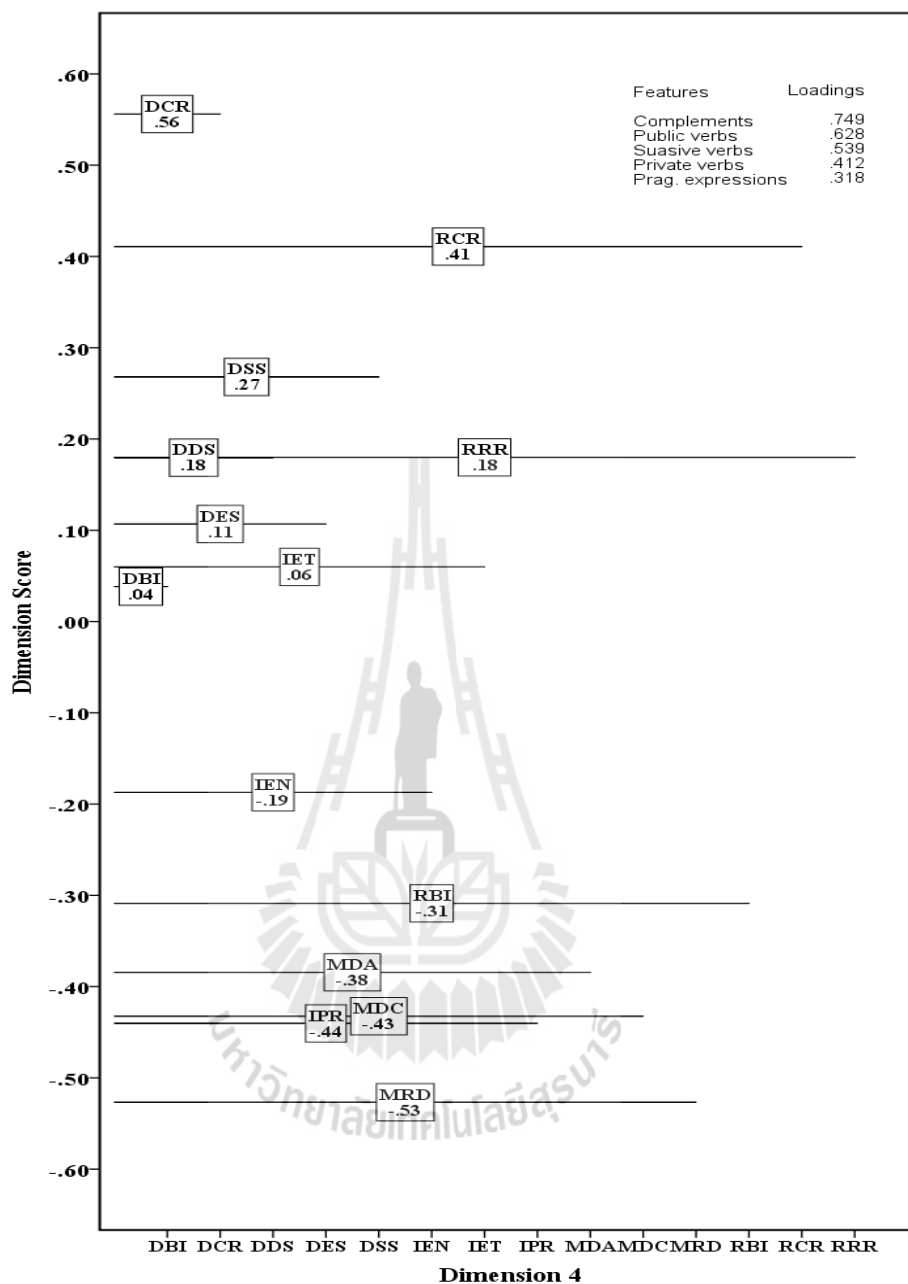
Figure 5.4 Moves along Dimension 4

As explained in 5.1.1.4, the monopolar characteristic of the dimension indicates that the moves on the positive end are, in varying degrees, marked by the positive features, and that in contrast, those linguistic markers are largely absent from the moves on the negative end.

At the extreme positive end, *Commenting on specific results (DCR)* of the discussion section is distantly separated from the other moves, indicating that it is much more marked by a focus on commentary than any other move. This is not surprising since DCR serves the purpose of commenting on results mostly by means of interpreting and explaining them.

> *10) The results of this study {suggest} [that] CEOs with MBAs have a positive influence on CEP. Even after accounting for firm characteristics and individual-level characteristics we still {found} [that] CEO MBA education resulted in higher levels of CEP. Moreover, our post-hoc analysis {revealed} no significant differences in results when MBA program rankings were {considered}. This is contrary to claims by critics [that] MBAs from top-tier programs may be associated with more meaningful outcomes due to differences such as candidate quality and program selectivity (Citation). Our results {imply} [that] consistency in MBA curricula across programs (Citation) leads to similar CEP-related effects, regardless of the school's ranking, which reinforces the claim [that] it is the MBA training itself making a difference*..*(LAE_09)(D4 Score = 1.86)*

In the above DCR move, two defining features are strongly present. The five private and suasive verbs (in {…}) report, interpret and explain propositions (in this case, results) expressed by five *that*-complements (with *that* in **[…]**), realizing the author's communicative purpose of commenting on results.

In contrast, *Describing research design (MRD)* at the negative extreme best represents the negative moves.

> *11) As mentioned above, the goal of this experiment was to assess the effect of students' quality expectations on their media-based learning achievement, and the moderating role of content relevance for this effect. Students' quality expectations were manipulated with the supply of either positive or negative quality information about the instructional medium they were about to study with. Furthermore, students' perceptions of content relevance were manipulated with either*

low or moderate relevance information. The reason for sparing the high relevance condition was that one of the two "no effect" conditions (i.e., low relevance) seemed sufficient for testing the moderating function of content relevance on the effect of quality information. Taken together, Experiment 1 used a 2 x 2 between subjects design, whereby the effects of the two experimental factors quality information and content relevance were observed with respect to students' final achievement outcome. . ...(LAI_09)(D4 Score = -.99)

The above is an MRD move text that describes the research design. As anticipated, the text is literally "poles apart" from the DCR move at the positive extreme. As can be seen, the features marking the DCR move is totally absent, except for the solitary *that*-complement (*that one…seemed sufficient..*), which is not associated with any public, private or suasive verb. This is so simply because the primary function of MRD is to describe the research design, not to comment on research results as does DCR.

Three distinct clusters of moves are observed on the dimension. With identical low means above zero, *Making deductions from the study (DDS)* of the discussion section and *Reporting specific results (RRR)* of the results section form a pair. They have a low level of presence of the defining features. This suggests that the two are significantly less commentary-oriented than the other moves situated above them on the dimension while significantly more so than those below them.

Another group on the positive end consists of three moves, viz., *Evaluating the study (DES)* and *Providing background information (DBI)* of the discussion section, and *Establishing a territory (IET)* of the introduction section. They are similar to one another while different from the other moves in the amount of

commentary they convey. In fact, with extremely low means, they are minimally marked on the dimension.

12) Students who are placed in lower-level math courses are at a curricular disadvantage when they start high school (Citation) and are less likely to fulfill course requirements for postsecondary education (Citation). Similarly, students who fail courses in their first year of high school are more vulnerable to dropping out of school (Citation). Although research has established that low-achieving middle school students are likely to have low academic outcomes in high school (Citation), this pattern of low achievement also depends in part on the changing context of the middle and high schools during the transition (Citation).....(SOE_01)(D4 Score = .05)

The above IET text is typical of the moves in the group with respect to this dimension. As can be seen, only the last sentence contains a *that*-complement clause, a feature that may mark an orientation towards commentary if it co-occurs with public, suasive, and private verbs and pragmatic expressions. As a result, this text is hardly relevant to commentary.

Likewise, with moderately low means on the negative pole, four moves form a homogeneous group characterized by a moderately low degree of irrelevance to commentary. They are *Presenting the present research (IPR)* of the introduction section, *Providing background information (RBI)* of the results section, *Describing data collection procedures (MDC)* and *Describing data analysis procedures (MDA)* of the methods section.

The remaining moves that do not group with other moves are significantly different from the other moves in the degree in which they are commentary-oriented or commentary-irrelevant. They are *Commenting on specific*

results (RCR) of the results section and *Summarizing the study (DSS)* of the discussion

section on the positive end, and *Establishing a niche (IEN)* of the introduction section

on the negative end.

5.1.2.5 Variation along Dimension 5

Figure 5.5 illustrates the relative standings of the moves on Dimension

5 labeled *Personal engagement vs. modified information*.

Making deductions from the study (DDS) sits the highest at the extreme

positive end of the dimension. As described in 4.1.5.4, the DDS move typically

suggests what to do in practice and/or in future research on the basis of the results of

the study. The following is a DDS move.

13) (We) noted that preservice teachers struggled with how {to teach} science at the preschool level. Preservice teachers currently \receive\ science teaching preparation in an integrated curriculum through /which/ (they) \are\ {to deliver} all instruction. (We) \recommend\ that when preschool teachers \engage\ their students in science activities (they) \label\ the activities as science as (they) would label reading and writing activities. <Doing> so would help (their) students get an initial awareness of science. Another recommendation \is\ {to refer} only to actual science investigations or content as "science" rather than use a broad definition /that/ incorporates many activities {to avoid} confusion regarding preschool science. The definition of science \may\ be broader in preschool (Citation) and should be more in line with elementary content. .... (RST_01)(D5 Score = 2.5)
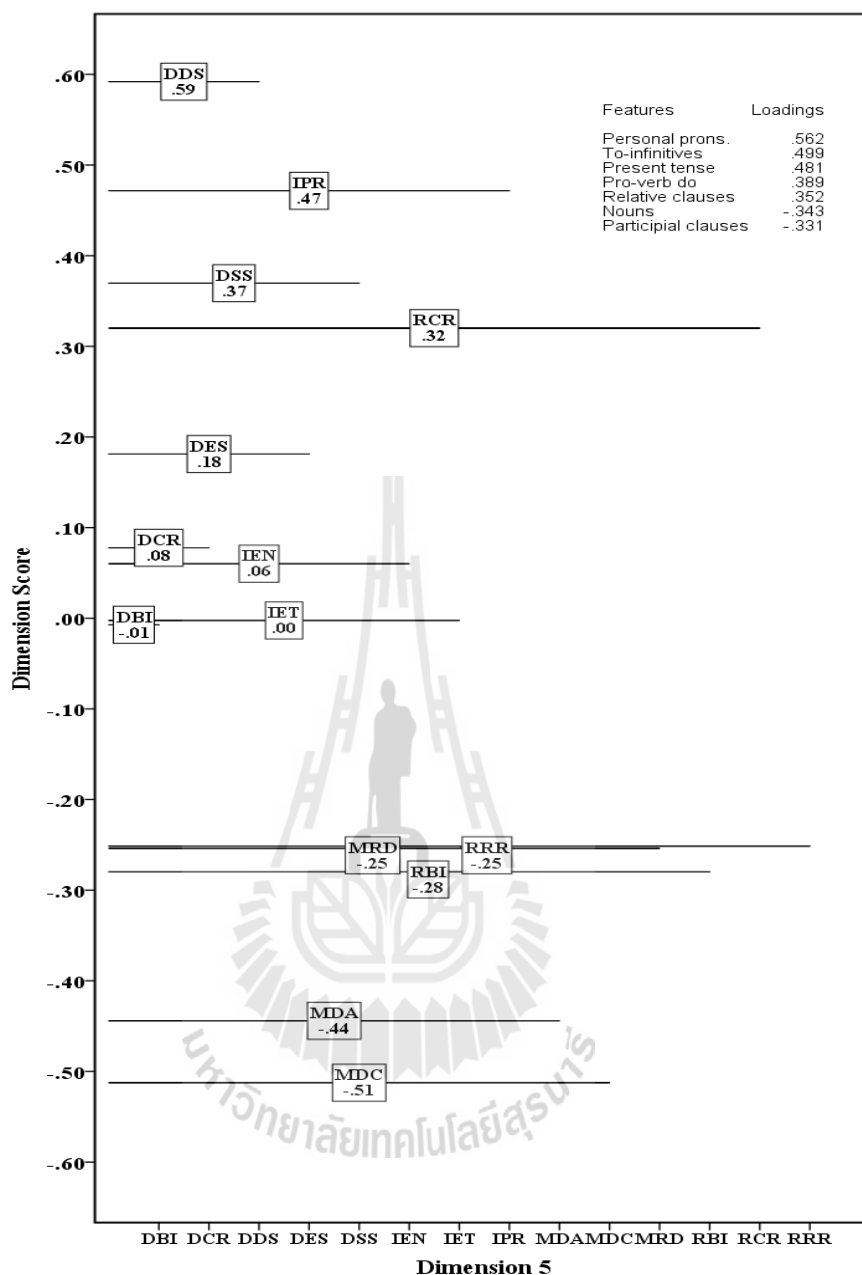
Figure 5.5 Moves along Dimension 5

In the text, all defining features of the positive end of the dimension have a pronounced presence: seven first and third person plural pronouns (in (…)), five present tense verbs (in \...\), four *to*-infinitives (in {…}), one pro-verb *do* (in <…>), and two relative clauses (introduced by relative pronouns in /…/). Together,

these features express a strong sense of personal engagement that involves "we," the authors, recommending what "they," preschool teachers, should do to improve practice. In engaging themselves in making recommendations, the authors obviously wish to engage the readers in what they think beneficial to do.

At the opposite extreme, *Describing data collection procedures (MDC)* pairs up with *Describing data analysis procedures (MDA)*, both of the results section. With close means which are the highest on this end, they only differ slightly in the degree in which they convey integrated information with nouns and participial clauses. Simultaneously, they are significantly different from the others. The following MDC move is as an example.

14) QOL was measured using the 31-item Quality of Life in Epilepsy scale, adapted from the more comprehensive 89-item scale and designed to address key epilepsy QOL issues. The QOLIE-31 contains seven subscales and one question addressing overall health; subscale concepts are emotional well-being, social function, energy/fatigue, cognitive function, seizure worry, medication effects and overall QOL. Scores are obtained by calculating a weighted average of the subscales using a scoring manual provided by the QOLIE-31 development group. The QOLIE-31 instrument has been reported to be both reliable and valid. The scale was found reliable for the present sample (a = 0.74-0.88 for subscales and a = 0.94 for the overall scale). ....(HER_05)(D5 Score = -1.40)

The text is densely populated with nouns modified by present participial or past participial phrases (e.g., *one question addressing overall health*; *a scoring manual provided by ...*) which condense information by reducing much space that would otherwise be needed if relative clauses were to substitute for them. Likewise, a couple of dangling present participial phrases further contribute to information

integration (e.g., *using the ...scale*). Clearly, a great deal of information is required to fulfill the communicative purpose of MDC moves, i.e., to describe data collection procedures in detail.

Four moves are close-knit around the dimension mean, indicating that they are mixed in orientation towards personal engagement and towards modified information. They are *Commenting on specific results (DCR)* and *Providing background information (DBI)* of the discussion section, and *Establishing a territory (IET)* and *Establishing a niche (IEN)* of the introduction section. IET best represents the group as it has a mean of zero.

15) Textbooks are central instructional media and should reflect broad educational emphases, and changes in [them], over time. [They] fall, organizationally, between {generalized educational policy agendas} and the actual instructional patterns [to be found] in classrooms. [They] are core features of the {intended curriculum}. ....*(IET_SOE_04)(D5 Score = 0)*

The IET text has a balanced presence of positive (three third person plural pronouns and one *to*-infinitive in [...]) and negative features (two nouns modified by past participles in {…}). The positive group neutralizes the negative group, resulting in a text unmarked either by personal engagement or by modified information. This is in consistence with the communicative purpose of this particular move, which is to establish a research territory by means of topic generalizations.

On the positive pole, *Summarizing the study (DSS)* of the discussion section and *Commenting on specific results (RCR)* of the results section form a noticeable pair. With moderately low means, they share not so strong a tendency

towards expressing personal engagement while significantly different in this respect from the other moves.

Still another group of four moves is situated on the negative end within a very short distance from the dimension mean. They are *Describing research design (MRD)* of the methods section, and *Providing background information (RBI)* and *Reporting specific results (RRR)* of the results section. Significantly different from MDA and MRD further down the negative pole, they are only slightly oriented towards modified information.

Neither *Presenting the present research (IPR)* nor *Evaluating the study (DES)* forms a group with other moves. They are marked by different levels of a personal engagement focus. The former, with the second highest positive mean, is moderately strong in this respect whereas the latter is only slightly marked due to a very low mean.

5.1.2.6 Variation along Dimension 6

Figure 5.6 shows the relations of the moves with respect to Dimension 6 labeled *Unsatisfactory status quo vs. research conduct*.

With an exceptionally high mean, *Establishing a niche (IEN)* of the introduction section sits high up on the positive pole of the dimension interpreted as representing the unsatisfactory status quo. As discussed earlier, while the present state of affairs is conveyed by present tense, present perfect aspect, existential sentences, and adverbial clauses, dissatisfaction with it is verbalized in negative words.
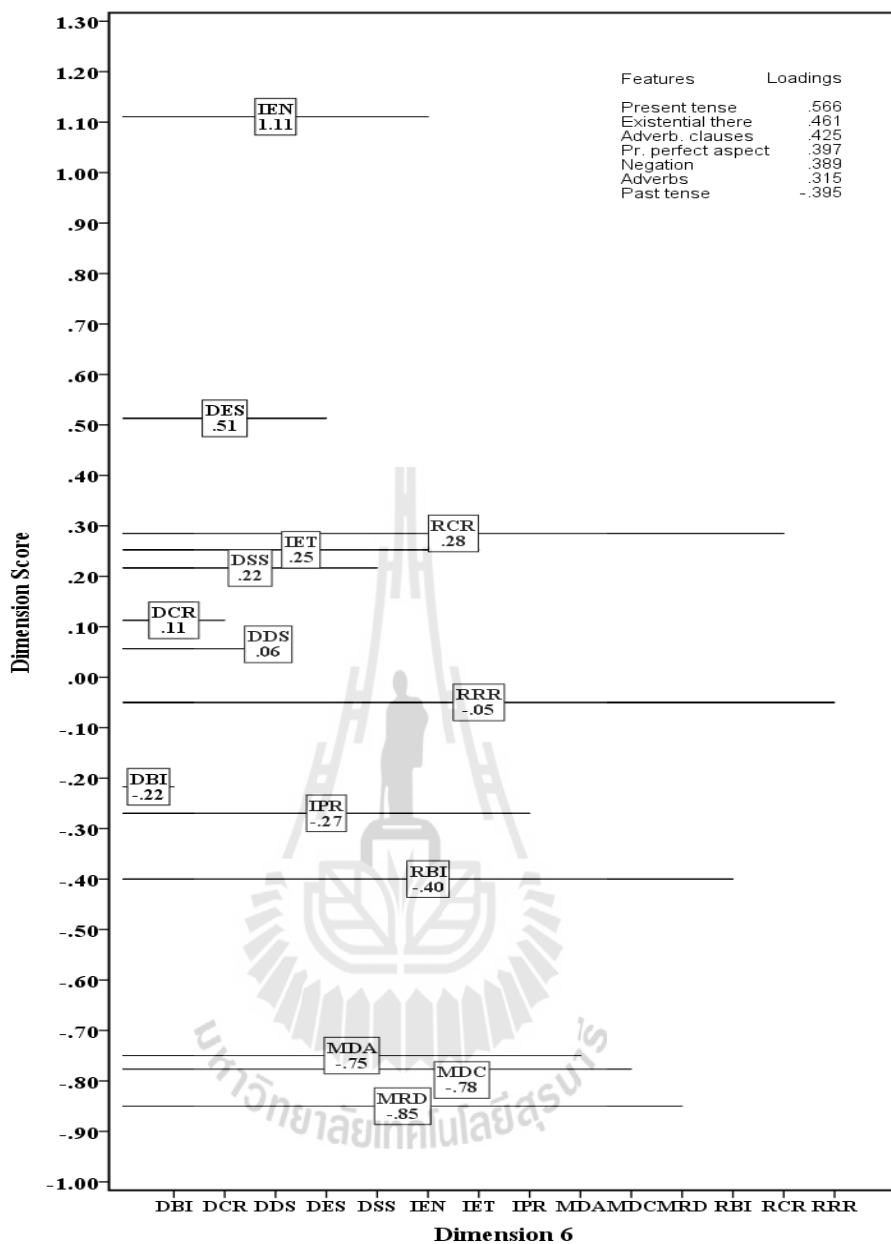
Figure 5.6 Moves along Dimension 6

        16) Although some studies have examined implicit corrective feedback in text-based SCMC (Citation), and other studies have looked at explicit corrective feedback in asynchronous CMC (Citation); no study to date has compared peer corrective feedback in text-based SCMC to asynchronous peer corrective feedback based on chat logs. .... (LLT_01)(D6 Score = 4.96)

        The above sample of an IEN move is rich in features that mark

dissatisfaction with research carried out so far in the relevant area: one adverbial clause of

concession (*Although* …), three instances of past perfect aspect (*have examined*; *have looked*; *has compared*), and most importantly a negative word (*no*). The propensity of the co-occurring features to express dissatisfaction with the status quo coincides with the communicative purpose of IEN which is to identify a gap in existing research.

Quite expectedly, occupying the farthest negative end is a cluster of three moves with remarkably high means: *Describing research design (MRD)*, *Describing data collection procedures (MDC)* and *Describing data analysis procedures (MDA)*. These moves, all belonging to the methods section, are far more strongly marked by a focus on research conduct than any other move or group of moves. Again, as shown earlier, this end is marked by past tense, which denotes past actions or situations and in addition, connotes temporal sequencing. The following MRD text with past tense verbs in […] captures the characteristics of this group of moves on this dimension.

> *17) We [employed] a primarily quantitative research design using a Web survey constructed by the research team. Most survey items [were] multiple choice or otherwise quantitative, but open-response items [were] also included to allow clarification. In mixed methods terminology, we [followed] an embedded (Citation) or concurrent nested (Citation) research design, in which the data [were] analyzed separately according to their respective traditions (quantitative or qualitative) and mixed during the analysis phase.* ....*(JEE_03)(D6 Score = -1.53)*

The text informs the reader what research design was employed, what instrument was used, and how data was analyzed. With no exception, all sentences are in past tense. Although the purpose of MRD is to outline the research design rather than narrate data collection or data analysis procedures in chronological order, sequencing can still be sensed in this particular text.

Conspicuously clustering around the mean of the dimension are *Reporting specific results (RRR)* of the results section, and *Making deductions from the study (DDS)* and *Commenting on specific results (DCR)* of the discussion section. These moves are the least marked due to their extremely low means. The RRR move text below best illustrates this characteristic.

> 18) *A Spearman rank correlation test, as shown in Table 3, revealed moderate and significant correlations between presence and arousal-sleepy quality, r_s (200) = 0.50, p < 0.001, and between presence and pleasant-unpleasant quality, r_s (200) = 0.55, p < 0.001. With regards to learning outcomes, only perception of learning is moderately correlated with presence, r_s (200) = 0.59, p < 0.001, the arousal-sleepy scale, r_s (200) = 0.40, p < 0.001, and with the pleasant-unpleasant scale, r_s (200) = 0.49, p < 0.001. However, none of the dependent measures correlates significantly with perceived difficulty, or with retention or transfer scores.* .... *(CAE_07)(D6 Score = .04)*

In this RRR text, both positive and negative features are present: two instances of a positive feature (present tense verbs *is* and *correlates*) and one instance of a negative feature (past tense verb *revealed*). However, due to the absence of other features to co-occur with them, they hardly convey any orientation towards either the unsatisfactory status quo or research conduct.

Three moves, namely, *Summarizing the study (DSS)*, *Establishing a territory (IET)*, and *Commenting on specific results (RCR)*, group together on the positive end with very low means, indicating that they are minimally marked by an orientation towards the unsatisfactory status quo.

Meanwhile, two moves form a noticeable pair on the negative end: *Providing background information (DBI)* of the discussion section and *Presenting the present research (IPR)* of the introduction section. Their low means suggest a weak tendency towards narrating research conduct.

Finally, neither *Evaluating the study (DES)* of the discussion section nor *Providing background knowledge (RBI)* of the results section groups together with any other move. The former, with a rather high mean on the positive end, is quite strongly marked by the unsatisfactory status quo whereas the latter, with a moderately high mean, are moderately oriented towards a focus on research conduct.

5.1.2.7 Variation along Dimension 7

Figure 5.7 shows the moves plotted along Dimension 7 representing frequent references to the present research on the positive end and large amounts of information about the past on the negative end. Similarities and differences among the moves are clear, based on the distances between individual moves or move groups.

Commenting on specific results (RCR) of the results section has the highest mean on the positive end. Thus, it exhibits the strongest orientation towards mentioning the research itself or any other aspect of the research. This basic communicative function is realized by demonstratives (*this*, *that*, *these*, *those*), coupled with prepositional phrases which pack large amounts of information expressed by nouns, as illustrated by the following text of an RCR move.
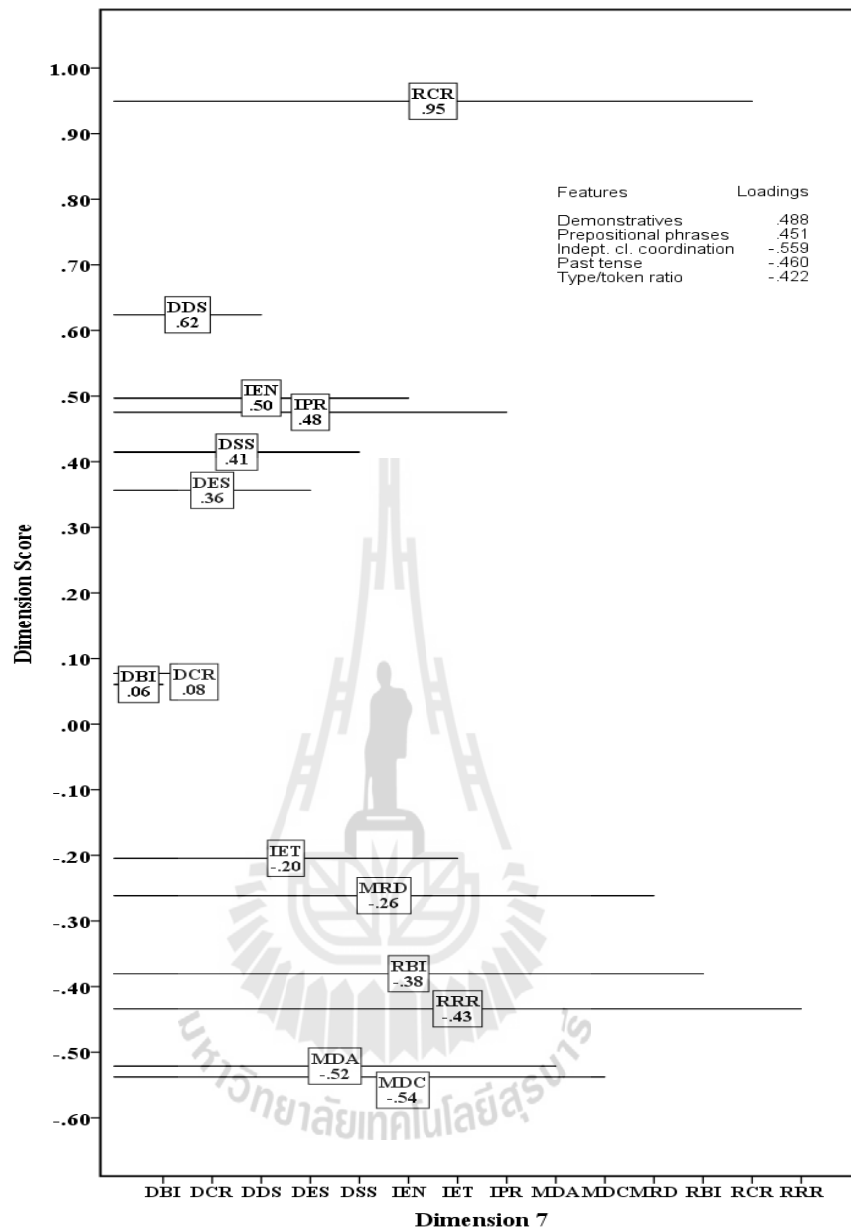
Figure 5.7 Moves along Dimension 7

19) While these correlations seem to suggest that participants who could not generate new hypotheses turned to the helpfiles after having conducted some exploratory experiments, the percentage of exploratory experiments} was too low to validate this pattern in scores from the logfiles ....(LAI_02)(D7 Score = 3.72)

In this RCR move, the author comments on some particular results by

making sense of and accounting for some results already presented. As can be seen,

three phrases, viz., *these correlations*, *this pattern in scores from logfiles* and *the percentage of exploratory experiments*, play a vital role in realizing this commentary. While the demonstratives in the first two (*these*, *this*) direct the reader's attention to the research outcomes, the prepositional phrases along with the nouns they modify in the second and third phrases provide concrete yet condensed information about the results being commented on.

At the other extreme of the dimension are two moves that pair up with moderately high means: *Describing data collection procedures (MDC)* and *Describing data analysis procedures (MDA)*, both of the methods section. As shown previously, this end indexes density of information relevant to the past. The manifest features are independent clause coordination (*and* as clause coordinator), high values of type/token ratio and past tense.

> 20) *During the 2008-2009 academic year, 40 students at a high school in the southeastern United States [participated] in this study. The students' mean age [was] 16.33 years (SD = .526) {and} 21 (52.5%) of them [were] female. Pretest scores [confirmed] that the majority of the participants [had] minimal knowledge of the history content, with 85 percent of them receiving the lowest score in terms of their conceptual understanding of the Regulator Movement...........................*
> *...(CAE_01)(D7 Score = -1.67)*

The MDC text above provides a large amount of information about the sample used in the research. The second sentence is comprised of two independent clauses, one providing information on age and one conveying information on gender. The coordinator *and* integrates the two pieces of information. Likewise, while a high type/token ratio (0.60) contributes to information density, four past tense verbs (in

{…}) define the information as relevant to the past. It should be noted that the above excerpt contains a demonstrative (*that*), an important feature marking the positive end. However, a single reference to the present research does not suffice to color the entire text with a focus on references to the present research.

Two moves are so close to the dimension mean that they are virtually unmarked on the dimension. They are *Providing background information (DBI)* and *Commenting on specific results (DCR)* of the discussion section. The following DBI text characterizes the moves on this dimension.

21) The current work {explored} an approach to encourage young women to pursue [careers in engineering] by targeting women's negative [beliefs about engineering] and their [lack of confidence in their own abilities] using computer-based anthropomorphic social models. Of particular interest {was} the [impact of the model's race and gender for its effectiveness] and whether matching the [race and gender of the participant] to the computer-based social model {increased} the [efficacy of this approach]. Drawing from the [literature on social models] (Citation), we {had} predicted that the agents that {were} more similar to the participants in race and gender {would} be the most effective in changing the participants' [beliefs about engineering]. .... (JEE_01)(D7 Score = .04)

The text performs the function of preparing the reader for a discussion of the results by restating the research objectives, methods, and hypothesis. Therefore, it is both necessary to refer to certain aspects of the research and provide the reader with some information to facilitate understanding of the discussion that follows. As can be seen, both positive and negative features abound in the text. Positive features include one demonstrative (*this*) and eight nouns modified by prepositional phrases (in […]); negative features are a high type/token ratio (0.64) and five past tense verbs

(in **{…}**). The result is a mixed text in that it does not seem to be biased towards either a focus on references to the research itself or towards a focus on providing information about the past.

On the positive end is another close-knit group of four moves: *Evaluating the study (DES)* and *Summarizing the study (DSS)* of the discussion section, and *Establishing a niche* (IEN) and *Presenting the present research (IPR)* of the introduction section. With low to moderately high means, the moves share a moderately strong likelihood to make references to certain aspects of the study. Their communicative purposes necessitate direct mentioning of the research itself or aspects of the research.

Additionally, two pairs of moves are visible on the negative end. One consists of *Providing background information (RBI)* and *Reporting specific results (RRR)* of the results section. Their moderately high means suggest that they have a moderately strong tendency towards a focus on information about the past. The other, comprising *Establishing a territory (IET)* of the introduction section and *Describing research design (MRD)* of the methods section, are only slightly marked by an informational focus due to very low means.

5.2 Discussion

Since the previous section has already interpreted the functional dimensions in association with the rhetorical moves, this section will only focus on the general characteristics of the dimensions.

The results of the multidimensional analysis of the ERC corpus show that education RAs have seven functional dimensions. Except for Dimension 4, all dimensions consist of two complementary poles, each marked by a set of co-occurring linguistic features. Each dimension serves as a parameter measuring how much each move makes use of the co-occurring linguistic markers. Measuring each move on all seven dimensions resulted in rich and accurate descriptions of the moves. The readings of the moves on the seven dimensions (i.e., their mean scores) indicate that some rhetorical moves use similar amounts of the linguistic features on one or more dimensions while using different amounts of the co-occurrence patterns of features marking other dimensions.

In contrast to Biber's inter-genre study, the present study is an intra-genre, inter-move investigation. As reviewed in Chapter 2, Biber (1988) investigated a range as broad as 23 registers/genres of spoken and written English. His purpose was to capture inter-genre linguistic variation. The six dimensions he identified serves to measure the ways the genres are similar or different. As his corpus covered a wide range of genres across speech and writing, the dimensions are seen as representative

of broad, fundamental communicative functions of the English language in general. In contrast, the present study focuses on a single academic genre of education research RAs, attempting to capture the linguistic variation among the rhetorical moves. In consequence, the dimensions uncovered tend to represent communicative functions that are less broad, less fundamental but more restricted than Biber's dimensions.

For example, Dimension 4 is restricted to making comments about propositions. As revealed by the move analysis, there is a *Commenting on specific results* move in both the results and discussion sections of education RAs. This move allows researchers to comment on propositions that express, for instance, research results or knowledge claims, in the results and discussion sections. It is not surprising to find that the *Commenting on specific results* move, shared by the discussion and results sections, is distinctly marked on this dimension.

The restrictedness of the dimensions is perhaps best reflected in Dimensions 6 and 7 that are apparently research-relevant. The communicative function underlying the positive end of Dimension 6 is to express dissatisfaction with the status quo. Coincidentally, the move that best represents this end of the dimension is *Establish a niche (IEN)* of the introduction section whose communicative purpose is to identify what has not been done satisfactorily in the relevant research area. The negative end of the dimension is strictly relevant to research conduct. The three methods section moves, namely, *Describing research design (MRD), Describing data collection procedures*

(MDC) and *Describing data analysis procedures (MDA),* are concentrated at the extreme of this end. This finding, which is in agreement with the global communicative purpose of the methods section, corroborates the move analysis results.

Likewise, both ends of Dimension 7 are directly associated with research. The positive end represents frequent references to the present research and the negative end has a focus on information about the past. *Commenting on specific results (RCR)* of the results section best represents the positive end as writers cannot possibly fulfill the communicative purpose of this move without mentioning the research itself or aspects of the research. Not surprisingly, *Describing data collection procedures (MDC)* and *Describing data analysis procedures (MDA)* of the methods section are distinctly marked on the negative end. As explained previously, there is a need for "thick" descriptions of the data collection and analysis procedures.

The restrictedness of the dimensions is highlighted by the finding that the co-occurring features often marking certain dimensions actually constitute a limited number of sentence patterns. For example, the linguistic features marking the positive end of Dimension 2 expressive of evaluative stance are 1) *Be* as main verb, 2) predicative, 3) pronoun *it*, 4) present tense, 5) pragmatic expressions and 6) complements (including *that*-verb complements, *that*-adjective complements, Wh-clauses, and infinitive complements). As indicated by the corpus data, these features often are not discrete entities; rather they form such basic sentence patterns as

1) It is adjective (e.g., *necessary*, *interesting*) to do something; 2) It is adjective (e.g., *surprising*, *unexpected*) that …; and 3) Wh-clause is adjective (e.g., *wrong*, *contradictory to*...). It is likely that the adjectives in these patterns are preceded with pragmatic expressions (e.g., *absolutely, almost, perhaps*). The sentence structures are apt ways of expressing evaluative stance. This finding conforms with that of Kanoksilapatham (2003, 2007), who observed in her corpus of biochemistry RAs a considerable number of "collocations" consisting of some co-occurring features.

The fact that four out of seven functional dimensions of the education RAs mark an information focus demonstrates that education RAs are heavily loaded with information. As revealed by the move analysis, education RAs are, to a great extent, characterized by the need to provide "thick" descriptions by means of various rhetorical moves and steps. With varying focuses, the information-relevant dimensions (Dimensions 1, 3, 5, and 7) fulfill specific functions related to the provision of rich information. For example, the focus on current information of Dimension 1 is well reflected in the *Establishing a niche (IEN)* and *Establishing a territory (IET)* moves that deal with the current situations of the relevant research field. In contrast, Dimension 7's focus on information about the past is best seen in the two important methods section moves: *Describing data collection procedures (MDC)* and *Describing data analysis procedures* (MDA). This information density reflected by several functional dimensions was also noted in biochemistry articles by Kanoksilapatham (2007).

Importantly, the results indicating significant differences among the moves along the dimensions demonstrate that the moves are characteristically distinguishable in terms of the degree in which they are marked by a given functional dimension underlying a set of co-occurring features. The inter-move linguistic variation along the dimensions lends support to the validity of move identification in the move analysis. The triangulation of move analysis and MD analysis has generated results that complement each other, providing a fuller and more valid description of the education RA genre. In particular, the results of move analysis aid in the interpretation of functional dimensions of variation identified by the MD analysis while those of the MD analysis add a linguistic dimension to the description of the rhetorical move structure generated by the move analysis.

In congruence with Kanoksilapatham (2003, 2007), the inter-move linguistic variation revealed by the MD analysis of education RAs highlights the complexity of academic writing. Rather than being uniform, unvarying discourse, academic writing can be quite variable both in terms of functional properties and linguistic realizations. Given its complex nature, a unidimensional, generalized description of even the small discourse unit of the rhetorical move is out of the question. Multiple perspectives are necessary to provide fuller and more accurate profiles of the different facets of rhetorical moves.

5.3 Summary

This chapter presented and discussed the results of the multidimensional analysis of the ERC corpus. First, the seven factors extracted with their co-occurring linguistic features were interpreted and labeled in terms of their basic communicative functions shared by the features. Then, the chapter reported the similarities and differences among the rhetorical moves as indexed by each of the seven functional dimensions. The results revealed that moves of the same section may be quite similar or significantly different in terms of the extent to which they are marked by the dimensions, and that moves of different sections may be similar on certain dimensions while significantly different on others. Subsequently, the results were discussed in light of relevant previous literature.

Chapter 6 will conclude the dissertation with a summary of the study, its implications and limitations, and recommendations for further research.

CHAPTER 6

CONCLUSIONS

This chapter concludes the dissertation. First, Section 6.1 summarizes the important aspects of the study including research objectives, methodology, and major findings. Then, Section 6.2 deals with theoretical and pedagogical implications of the study whereas Section 6.3 outlines its limitations. Finally, in Section 6.4, recommendations for further research are made.

6.1 Summary of the Study

This study was built on a rich tradition of research on the RA genre to investigate both the rhetorical move structure of education RAs and systematic linguistic variation among the moves. There were two underlying reasons for the study. For one, novice NNS education researchers, who constitute a huge population, often find it difficult to write up their research for international publication as required by career advancement. Their greatest difficulty is typically related to insufficient knowledge of the rhetorical structure of the RA genre and to their unsatisfactory linguistic competence. For the other, comprehensive and accurate descriptions of the education research article genre are currently unavailable despite the insights the huge amount of previous research has provided into the RA genre in general.

Table 6.1 Summary of Results

Rhetorical Moves		Status	Dimensions on which moves are most salient	Linguistic features on dimensions
IET	**Establishing a territory**	Obl	1, 6, -7	**Dimension 1**
IET1	Presenting a vignette	Unc		Mean word length
IET2	Claiming centrality	Con		Present tense
IET3	Making topic generalizations	Con		Nominalizations &
IET4	Reviewing items of previous research	Con		gerunds
IET5	Defining terms and concepts	Opt		Attributive adjectives
IEN	**Establishing a niche**	Con	6, 1, 7	Phrasal coordination
IEN1	(one or more of A-E)	Con		
IEN1	Counter-claiming	Opt		-Past tense
IEN1	Indicating a gap	Con		....
IEN1	Raising a question	Opt		
IEN1	Continuing a tradition	Opt		**Dimension 2**
IEN1	Postulating hypotheses	Unc		*Be* as main verb
IEN2	Providing positive justifications	Opt		Predicative adjectives
IDM	**Describing the conceptual model**	**Opt**		Pronoun *it*
IDM1	Introducing the model	Opt	Excluded	Present tense
IDM2	Identifying components	Opt		Pragmatic expressions
IDM3	Describing components' relationships	Opt		Complements
IPR	**Presenting the present research**	Obl	7, 5, 1	-Past tense
IPR1	Announcing the present research	Con		....
IPR2	Announcing principle findings	Opt		
IPR3	Indicating RA structure	Opt		**Dimension 3**
IPR4	Claiming significance of the research	Opt		Dispreferred forms
IPR5	Verifying compliance with ethical standards	Unc		Modals
IPR6	Describing research context	Opt		Passives
MRD	**Describing research design**	**Opt**	-6, -2, -4	Adverbs
MDC	**Describing data collection procedures**	**Obl**	-1, -6, -2*	Pronoun *it*
MDC1	Describing research context	Opt		*to*-infinitives
MDC2	Describing the sample	Con		-Nouns
MDC3	Describing instruments	Opt		-Prepositional phrases
MDC4	Elaborating on data collection procedures	Opt		....
MDC5	Justifying data collection procedures	Opt		**Dimension 4**
MDC6	Verifying compliance with ethical standards	Opt		Complements
				Public verbs
				Suasive verbs
				Private verbs
				Pragmatic expressions

Note. * "–" indicates positive features on the dimension

Table 6.2 Summary of Results (Cont'd)

Rhetorical Moves		Status	Dimensions on which moves are most salient	Linguistic features on dimensions
MDA	**Describing data analysis procedures**	**Con**	-6, -1, -2	**Dimension 5** Personal pronouns *to*-infinitives Present tense Pro-verb *do* Relative clauses
MDA1	Recounting data analysis procedures	Opt		
MDA2	Justifying data analysis procedures	Opt		
MDA3	Establishing inter-coder reliability	Opt		
RBI	**Providing background information**	**Opt**	-3, -2, -1	-Nouns -Participial clauses ....
RRR	**Reporting specific results**	**Obl**	-1, -7, -3	
RRR1	Preparing for specific results presentation	Con		
RRR2	Presenting results	Obl		
RRR3	Illustrating with examples or excerpts	Opt		**Dimension 6** Present tense Existential *there* Adverbial clauses Present perfect aspect Negation Adverbs
RCR	**Commenting on specific results**	**Con**	7, 2, 3	
RSR	**Summarizing results**	**Opt**	Excluded	
DBI	**Providing background information**	**Con**	-2, 1, -6	
DCR	**Commenting on specific results**	**Obl**	4, 2, 6	
DCR1	Restating hypotheses or research questions	Opt		
DCR2	Restating results	Con		-Past tense ....
DCR3	Interpreting results	Con		
DCR4	Comparing results with literature	Con		
DCR5	Accounting for results	Con		**Dimension 7** Demonstratives Prepositional phrases -Independent clause coordination
DCR6	Illustrating with examples or excerpts	Opt		
DCR7	Recommending future research	Opt		
DES	**Evaluating the study**	**Con**	2, 6, 3	
DES1	Indicating significance	Opt		
DES2	Acknowledging limitations	Con		-Past tense -Type/token ratio
DDS	**Making deductions from the study**	**Con**	3, 7, 5	
DDS1	Drawing implications or making suggestions	Con		
DDS2	Recommending future research	Con		
DSS	**Summarizing the study**	**Opt**	1, 7, 2	

Swales' (1981, 1990, 2004) move analysis and Biber's MD analysis (1986, 1988) provided appropriate frameworks within which this study was conducted. Using as primary data the self-constructed, representative corpus of education research articles (ERC), a move analysis was conducted to identify the rhetorical moves and their generic structure in the corpus while an MD analysis was performed to uncover the functional dimensions underlying the co-occurring linguistic features that mark the different rhetorical moves.

Tables 6.1 and 6.2 present a summary of the major findings of the study. They display the moves and steps identified in their most likely order, and their status as well; they also display the dimensions uncovered and the dimensions on which the moves are most salient (Note again that two moves are not reflected on the dimensions because they were excluded from the MD analysis due to insufficient number of observations).

In answer to Research Questions 1 and 2, the move analysis results show that education researchers have a large repertoire of 16 rhetorical moves to choose from when writing research articles. These moves and steps are used in varied configurations and they vary in terms of importance, ranging from obligatory to unconventional. Apart from moves and steps common to all fields, education researchers also use a few unique moves and steps such as *Describing the conceptual model* move in their introduction sections and the *Verifying compliance with ethical standards* step in their methods sections. Moreover, they tend to use a few moves or steps differently from researchers in other fields. The specific vs. general distinction regarding the use of the *Announcing the present research* step is an example. In addition, education researchers frequently recycle their moves in all the sections in varying degrees. They do not only recycle individual moves but also repeat move patterns in larger cycles.

Besides structural complexity, rich description is another characteristic of the education RA genre. Many move and step labels feature the verb "describe," particularly for the methods section. The research design, data collection procedures including the research context, sample, instruments, and data analysis procedures are usually described in remarkable detail. Even the introduction section occasionally contains descriptive details, as in the vignettes provided to highlight a situation, or when introducing a conceptual framework.

They present study ascribes much of the structural complexity and detailed description to the qualitative nature characteristic of much of education research. The diversity of education research makes it necessary for education research writers to employ a large number of moves and steps for different purposes, and often to recycle them or their combinations as many times as there are, for example, hypotheses, research questions and results. Similarly, the need to establish reliability and validity and the need to facilitate the reader's understanding of their results give rise to thick descriptions of various aspects of their research.

In answer to Research Questions 3 and 4, the MD analysis uncovered seven functional dimensions, each underlying a set of co-occurring linguistic features. To recapitulate, these dimensions are 1) *Current information vs. procedural concerns*, 2) *Evaluative stance vs. past actions or states*, 3) *Logical probability vs. integrated information*, 4) *Commentary*, 5) *Personal engagement vs. modified information*, 6)

Unsatisfactory status quo vs. research conduct, and 7) *References to present research vs. Information about the past.*

In terms of the extent to which the moves are marked by the co-occurring linguistic features on each functional dimension, both similarities and differences were found among the moves on all dimensions. Moves that group together with other moves with close dimension mean scores are similar while those some distance away from others differ from them. Similar moves on each dimension are summarized as follows. The rest are significantly different.

Dimension 1: 1) IEN, DSS and IET; 2) MDC, MDA and RRR; 3) DCR, DES, and RCR; and 4) DDS and IPR.

Dimension 2: 1) RCR and DES); 2) DDS, DCR, DSS and IEN; 3) MDC and MDA; and 4) IPR, DBI and RBI.

Dimension 3: 1) DES, RCR, and IEN; and 2) IET, IPR, MRD, MDA.

Dimension 4: 1) DDS and RRR; 2) DES, DBI and IET; and 3) IPR, RBI, MDC and MDA.

Dimension 5: 1) MDC and MDA; 2) DCR, DBI, IET and IEN; 3) DSS and RCR; and 4) MRD, RBI, and RRR.

Dimension 6: 1) MRD, MDC and MDA; 2) RRR, DSS, DDDS and DCR; 3) DSS, IET, RCR; 4) DBI and IPR; and 5) DBI and IPR.

Dimension 7: 1) MDC and MDA; 2) DBI and DCR; 3) DES, DSS, IEN and IPR; 3) RBI and RRR; and 4) IET and MRD.

The results of the MD analysis further demonstrate the complexity of the education RA genre by revealing linguistic variation among the rhetorical moves. While moves can be similar on certain dimensions, they can be quite different on others. In other words, the rhetorical moves in education RAs are multi-faceted in terms of functional properties and linguistic realizations, requiring multiple perspectives for comprehensive description.

In summary, the synergy of move analysis and MD analysis resulted in enriched characterization of the education research article genre. The two-level structural models for each of the four sections and the readings of each move on each dimension underlying the co-occurrence patterns of linguistic features represent fuller and more accurate delineations of the genre.

6.2 Theoretical and Pedagogical Implications

This study has both theoretical and pedagogical relevance. Theoretically, it helps clarify the relationship between function and form. Genre analysis distinguishes itself from other approaches to text by using communicative purpose as the primary parameter for classifying discourses. With regards to move analysis, ostensibly no analyst disagrees that just as the overall communicative purpose shapes the genre, the

local communicative purposes mark the rhetorical moves. However, despite the general agreement on communicative purpose (i.e., function) as determinant of genre and its constituent rhetorical moves, there seems to be no actual consensus on the extent to which the local communicative purposes should be relied on in classifying rhetorical moves. In practice, function-based (Kwan, 2006; Pho, 2008b; B. Zhang et al., 2012), form-based (e.g., K. Anderson & Maclean, 1997; Lim, 2006) and function and form based approaches (e.g., Kanoksilapatham, 2007; Nwogu, 1997; Swales, 1990) have existed side by side. The relationship between function and form is thus quite blurry. By adhering to the notion of purpose determining form, the present study identified moves solely by their functions irrespective of linguistic forms. The subsequent MD analysis empirically uncovered sets of co-occurring linguistic features (i.e., the seven dimensions), each of which makes a unique contribution to the realization of the communicative purpose of each rhetorical move. The basic communicative functions of those sets of features, in conjunction, realize the communicative purposes of the rhetorical moves. Thus, the rhetorical moves identified solely on the basis of communicative function were validated by the results of the MD analysis of linguistic features. In sum, the results of the study contribute to genre theory by demonstrating that communicative purpose is not slippery but quite operational in identifying rhetorical moves.

The strongest motive for genre research is the input it provides for ESP/EAP instruction. The pedagogical implications of this study are found in the potential values of the results informing syllabus design and materials development for courses aimed at helping novice NNS education researchers, pre-service or in-service, to write research articles for publication. Results about the rhetorical moves of education research articles and linguistic realizations integrated into the syllabus and course materials can ensure participants' awareness and understanding of the generic conventions of the discourse community of education research, and provide them with useful guidelines for structuring their articles.

In particular, due attention should be paid to variation between the specific genre of the education research article and the general research article genre. Although the rhetorical structure of education research articles was found to be in general conformity with research articles in other disciplines, some moves and steps are unique to education research articles. Since these idiosyncrasies often arise from the qualitative nature characteristic of much of education research, it is imperative to give them focal attention along with explanations in relation to qualitative research.

For example, the present study identified a new, stand-alone *Describing the conceptual framework* move and discriminates it from descriptions of conceptual frameworks intertwined with literature being reviewed. It makes much sense to prioritize teaching how to organize a conceptual framework in one clearly defined block that is characterized by reduced complexity and enhanced manageability. The

more difficult mode of conceptual framework organization should wait until after the course participants have mastered the free-standing conceptual framework.

Another example of rhetorical moves that should be included in instruction is the *Verifying compliance with ethical standards* step quite commonly employed by education writers. Education research frequently involves human subjects. Whether it is required by the journals or not, insofar as the study is potentially threatening to human subjects, including a verification statement in the research paper is recommended. Therefore, it becomes worthwhile to teach the writing of verification statements. This is especially true if participants of the course are from countries without an established mechanism, like the Institutional Review Board (IRB) in American institutions that reviews grant proposals and makes sure that they comply with ethical and legal requirements.

Still another example of variation that merits pedagogical consideration is the remarkably common specific type of *Announcing the present research* step. This study uniquely distinguishes the specific and general *Announcing the present research* step of the *Presenting the present research* move of the introduction section. Unlike the general type described in previous research as providing an outlined, general preview of the research towards the end of the introduction, the specific type mentions one particular aspect of the current study once a specific research gap was identified. Its inclusion in course materials seems necessary.

In light of Lave and Wenger's (1991) notion of "legitimate peripheral

participation," Kanoksilapatham (2011) argues that novice NNS writers should be encouraged to strike a balance between conforming to norms of the discourse community and maintaining their "primary ownership" of their manuscripts by retaining some of their idiosyncratic features. This, she adds, would contribute to research quality enhancement. However, to avoid the disastrous fate of their articles being rejected, it is the researcher's view that novice writers should be discouraged from flouting any generic convention. In the fierce competition for international publication, novice writers should play safe and always try to follow the models based on ample empirical evidence. Teachers should always keep in mind that the primary goal of ESP or EAP courses is to help learners acquire what Bhatia (2004) called "generic competence," that is, "the ability to identify, construct, interpret, and successfully exploit a specific repertoire of professional disciplinary or workplace genres to participate in the daily activities and to achieve the goals of a specific academic/professional community" (p.145).

Novice NNS writers are troubled not only rhetorically but also linguistically. Therefore, teaching the move structure alone is far from enough. Tables 6.1 and 6.2 provide a useful tool for teaching the linguistic devices typically used to communicate the purposes of the moves. The tables together provide all seven sets of co-occurring linguistic markers of the functional dimensions. For each rhetorical move, the three dimensions on which the move is the most salient are provided. They are arranged in the

order of the extent to which the move makes use of the three sets of linguistic features corresponding to the dimension numbers. While teaching the moves, a series of form-focused mini lessons can be conducted to first arouse the course participants' awareness of the linguistic devices and then provide them with opportunities for practice. More effective learning would occur if some well-chosen move texts from authentic education RAs are made available to the course participants. For instance, concordance lines generated from the ERC-M corpus would provide valuable reading for the learners.

6.3 Limitations of the Study

One major drawback related to the analysis of rhetorical moves is the limited generalizability inherent in all qualitative inquiries. First, the representativity of the corpus of 120 education research articles is not undisputable although it is among the biggest among the corpora used so far in studies of research articles. Second, the results are inevitably tinted with some degree of subjectivity although efforts were made to keep it at a minimum, for example, by involving a member of the relevant discourse community as inter-coder who, besides coding a quarter of the texts for reliability, provided much insight into the community's discursive practices and contributed immensely to the rigor of the research. Third, because it was impossible to get into close contact with the authors of the articles, no informant interview was conducted on any of them for more insights into their discoursive practices.

Another limitation was that the scope of the study did not allow for separate treatment of quantitative and qualitative research articles. Although the majority of the empirical education research articles in the ERC corpus were of a quantitative nature, a considerable portion had qualitative characteristics due to a mixed methods research design. In consequence, the relatively low frequencies of a few moves and steps unique to qualitative research might have obscured their importance.

Regarding the MD analysis of linguistic features, the total variance explained by the seven dimensions was not particularly high although it was slightly higher than Kanoksilapatham's (2007) figure. According to Biber (1995), MD analyses of co-occurrence linguistic patterns typically account for about 50 percent of the total variance, the remaining variation patterns of individual features being associated with less fundamental communicative functions. It is only a conjecture that the rather low percentage of total variance explained was due to the relative homogeneity of the education research article genre as compared to Biber's (1988, 1995) heterogeneous range of text varieties.

6.4 Recommendations for Further Research

This study has identified some new moves and steps employed by education researchers accountable by the qualitative research nature. Yet, to date, none of the previous genre studies has investigated the qualitative research article genre in its own

right. The vital role that qualitative methods and mixed methods play in social sciences and humanities research calls for investigations into the rhetorical move structure of the qualitative research article. Perhaps, besides investigations into qualitative research articles in general, a series of studies are necessary to capture the rhetorical organizations of research articles in different fields and disciplines.

Due attention might also be paid to the highly frequent clause-level structural patterns. The results of the MD analysis indicate that co-occurring features often constitute a limited number of larger, clause-level structural patterns. However, up to date, investigations into linguistic features have been restricted to units smaller than the clause, such as word classes and lexical bundles. Results from clause-level linguistic investigations in relation to the rhetorical move structure should be pedagogically beneficial. In the English language, clauses are constructed on a finite number of patterns. Identifying the frequent ones and then characterizing them by their communicative functions would provide much insight into different genres or registers as well as the English language as a whole.

It might also be interesting to see how the moves compare to Biber's registers along Biber's dimensions such as academic prose, romantic fiction, interviews, broadcasts and press editorials. This can be done by plotting the moves along Biber's dimensions and see what register each move is closest to on each dimension. Since the study examined the same 67 linguistic features that Biber investigated, it is

methodologically possible to plot the mean score of the moves on each of Biber's dimensions and examine their positions relative to the registers on the dimensions.

In conclusion, the present study represents a useful attempt to investigate the education research article genre from both a macro and micro perspective. By successfully answering the questions it asked, it has demonstrated the power of the combination of move analysis and MD analysis in capturing the rhetorical move structure of the education research article genre and the linguistic variation among the moves. Yet, it also raises some issues to be addressed by future research.
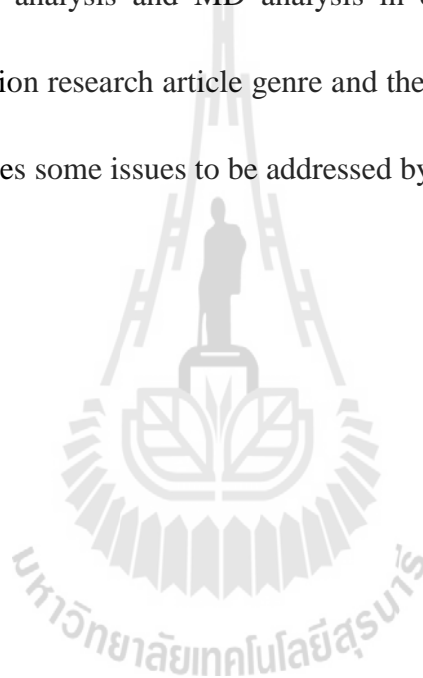
REFERENCES
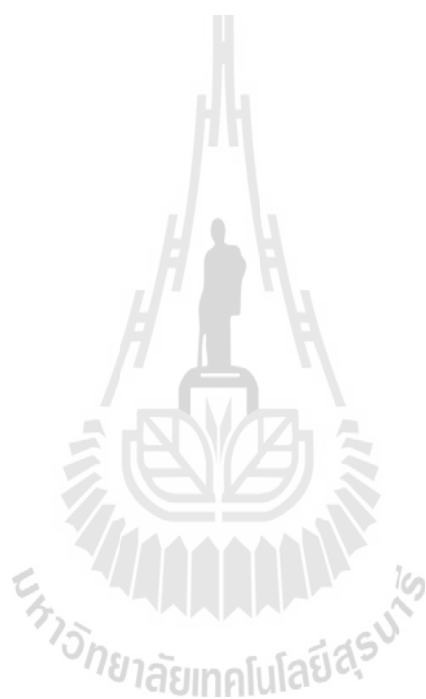
REFERENCES

Abdollahzadeh, E. (2011). Poring over the findings: Interpersonal authorial engagement in applied linguistics papers. *Journal of Pragmatics, 43*(1), 288-297.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington DC: American Psychological Association.

Amirian, Z., Kassaian, Z., & Tavakoli, M. (2008). Genre analysis: An investigation of the discussion sections of applied linguistics research articles. *The Asian ESP Journal, 4*(1), 39-63.

Amnuai, W., & Wannaruk, A. (2013a). An analysis of moves in introductions in international and Thai journal research articles. *PASSA, 45*, 61-60.

Amnuai, W., & Wannaruk, A. (2013b). Investigating move structure of English applied linguistics research article discussions published in international and Thai journals. *English for Language Teaching, 6*(2), 1-13.

Anderson, G. (1998). *Fundamentals of educational research* (2nd ed.). Philadelphia, PA: The Palmer Press.

Anderson, K., & Maclean, J. (1997). A genre analysis study of 80 medical abstracts. *Edinburgh working papers in applied linguistics, 8*, 1-23.

Anthony, L. (1999). Writing research article introductions in software engineering: How accurate is a standard model? *Professional Communication, IEEE Transactions on, 42*(1), 38-46.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/

Balocco, A. E. A. (2000). Who's afraid of literature? Rhetorical routines in literary research articles. *Especialist, 21*(2), 207-223.

Baratta, A. M. (2009). Revealing stance through passive voice. *Journal of Pragmatics, 41*(7), 1406-1421.

Bassey, M. (1999). *Case studies in educational settings.* Buckingham: Open University Press.

Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in language teaching. *Journal of English for Academic Purposes, 8*(4), 241-251.

Basturkmen, H. (2012). A genre-based investigation of discussion sections of research articles in dentistry and disciplinary variation. *Journal of English for Academic Purposes, 11*(2), 134-144.

Bawarshi, A. S., & Reiff, M. J. (2010). *Genre: An introduction to history, theory, research, and pedagogy.* West Lafayette, IN: Parlor Press.

Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bhatia, V. (1993). *Analysing genre: Language use in professional settings*. New York: Longman.

Bhatia, V. (2002). Applied genre analysis: A multi-perspective model. *IBÉRICA, 4*, 3-19.

Bhatia, V. (2004). *Worlds of written discourse: A genre-based view*. London: Contiunuum.

Biber, D. (1985). Investigating macroscopic textual variation through multifeature/multidimensional analysis. *Linguistics, 23*, 337-360.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 384-414.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing, 5*(4), 257-269.

Biber, D. (1993). Co-occurrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. *Computational Linguistics, 19*(3), 531-538.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Biber, D. (2008). Dimensions of variation in converstation. In B. K. Bhatia, J. Flowerdew & R. H. Jones (Eds.), *Advances in discourse studies* (pp. 100-114). London: Routledge.

Biber, D. (2009). Multi-dimensional approaches. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An international handbook* (Vol. 2, pp. 822-855). Berlin, Germany: Walter de Gruyter GmbH & Co.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins Publishing Company.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2003). The authors respond: Strengths and goals of multidimensional analysis. *RESEARCH ISSUES*, 151.

Biber, D., & Finegan, E. (1994). Intra-textual variation within medical research articles. In N. Oostidijk & P. d. Haan (Eds.), *Corpus-based research into language: In honor of Jan Aarts* (pp. 201-221). Amsterdam: Rodopi.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Beijing: Foreign Language Teaching and Research Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English*. Essex, England: Pearson.

Bondi, M. (2008). Emphatics in academic discourse Integrating corpus and discourse tools in the study of cross-disciplinary variation. In A. Ädel & R. Reppen (Eds.), *Corpora and discourse: The challenges of different settings* (pp. 31-55). Amsterdam: John Benjamins.

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes, 13*(1), 47-59.

Bruce, I. (2008). Cognitive genre structures in Methods sections of research articles: A corpus study. *Journal of English for Academic Purposes, 7*(1), 38-54.

Bruce, I. (2009). Results sections in sociology and organic chemistry articles: A genre analysis. *English for Specific Purposes, 28*, 105-124.

Brusaw, T., Alred, G. J., & W. E. Oliu. (1993). *Handbook of technical writing*. New York: St. Martin's.

Burrough-Boenisch, J. (2003). Shapers of published NNS research articles. *Journal of Second Language Writing, 12*(3), 223-243.

Cao, Y., & Xiao, R. (2013). A multi-dimensional contrastive study of English abstracts

by native and non-native writers. *Corpora, 8*(2), 209-234.

Cargill, M., & O'Connor, P. (2006). Developing Chinese scientists' skills for publishing

in English: Evaluating collaborating-colleague workshops based on genre

analysis. *Journal of English for Academic Purposes, 5*(3), 207-221.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral*

Research, I, 245-276.

Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral

literature. In D. Tannen (Ed.), *Spoken and written language: exploring orality*

and literacy (pp. 35-54). Norwood, N.J.: Ablex.

Chafe, W. L. (1985). Linguistic differences produced by differences between speaking

and writing. In D. R. Olson, N. Torrance & A. Hildyard (Eds.), *Literature,*

language, and learning: the nature and consequence of reading and writing (pp.

105-123). Cambridge: Cambridge University Press.

Chang, C.-F., & Kuo, C.-H. (2011). A corpus-based approach to online materials

development for writing research articles. *English for Specific Purposes, 30*(3),

222-234.

Cho, S. (2004). Challenges of entering discourse communities through publishing in

English: Perspectives of nonnative-speaking doctoral students in the United

States of America. *Journal of Language Identity and Education, 3*(1), 47-72.

Connors, R. J. (1981). The rise and fall of the modes of discourse. *College Composition and Communication, 32*, 444-455.

Conrad, S., & Biber, D. (2001). *Variation in English: Multi-dimensional studies*. Harlow, England: Longman.

Cooper, C. (1985). *Aspects of article introductions in IEEE publications.* (Unpublished M. Sc. thesis), University of Aston, UK.

Corbett, J. (2006). Genre and genre analysis. In B. Keith (Ed.), *Encyclopedia of language and linguistics* (2nd ed., pp. 26-32). Oxford: Elsevier.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*(1), 43-57.

Crookes, G. (1986). Towards a validated analysis of scientific text structure. *Applied Linguistics, 7*(1), 57-70.

Curry, M. J., & Lillis, T. (2004). Multilingual scholars and the imperative to publish in English: Negotiating interests, demands, and rewards. *TESOL Quarterly, 38*(4), 663-688.

Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics, 36*(10), 1807-1825.

Day, R. A. (1977). How to write a scientific paper. *IEEE, vol. PC-20*, 32-37.

Day, R. A., & Gastel, B. (2006). *How to write and publish a scientific paper* (6th ed.). Cambridge: Cambridge University Press.

Del Saz-Rubio, M. M. (2011). A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes, 30*(4), 258-271.

DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

ElMalik, A. T., & Nesi, H. (2008). Publishing research in a second language: The case of Sudanese contributors to international medical journals. *Journal of English for Academic Purposes, 7*(2), 87-96.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Fakhri, A. (2004). Rhetorical properties of Arabic research article introductions. *Journal of Pragmatics, 36*(6), 1119-1138.

Feng, H. (2006). A corpus-based study of research grant proposal abstracts. *Perspectives: Working Papers in English and Communication, 17*(1), 1-24.

Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing, 8*(3), 243-264.

Flowerdew, J. (2001). Attitudes of journal editors to nonnative speaker contributions. *TESOL Quarterly, 35*(1), 121-150.

Flowerdew, J., & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes, 29*(2), 78-93.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11-23). Amsterdam: John Benjamins.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*(3), 321-332.

Gavioli, L. (2005). *Exploring corpora for ESP learning* (Vol. 21). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Getkham, K. (2010). *A Corpus-based study of applied linguistics research articles: A multidimensional analysis*. Paper presented at the CILC 2nd International conference on Corpus linguistics, University of A Coruna, Spain.

Gillaerts, P., & van de Velde, F. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes, 9*(2), 128-139.

Gledhill, C. (1995). Collocation and genre analysis: The phraseology of grammatical items in cancer research abstracts and articles. *Zeitschrift für Anglistik und Amerikanistik: A quarterly of language, literature and culture, 1*(1), 11-29.

Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes, 19*(2), 115-135.

Goyvaerts, J. (2015). EditPad Light (Version 7.3.8) [Computer Software]: Just Great Software Co. Ltd.

Group, T. F. (2011). Structured abstracts: guide for authors. Retrieved Nov. 12, 2011, from http://www.tandf.co.uk/journals/authors/rereabstracts.asp

Gunawardena, C. (1989). Biochemistry research articles: Present perfect of biology and biochemistry. *English for Specific Purposes, 8*(3), 265-273.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). Oxford: Oxford University Press.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesin in English*. London: Longman.

Halliday, M. A. K., & Hasan, R. (1989). *Language, context and text: Aspects of language in a social semiotic perspective*. Oxford: Oxford University Press.

Harmon, J. E. (1989). The structure of scientific and engineering papers: A historical perspective. *IEEE, 32*, 132-137.

Hasan, R. (1989). The structure of a text. In M. A. K. Halliday & R. Hasan (Eds.), *Language, context and text: Aspects of language in a social semiotic perspective* (pp. 52-69). Oxford: Oxford University Press.

Henry, A., & Roseberry, R. L. (1999). Raising awareness of the generic structure and linguistic features of essay introductions *Language Awareness, 8*(3/4), 190-200.

Herrington, A., & Moran, C. (2005). The idea of genre in theory and practice: An overview of the work in genre in the fields of composition and rhetoric and new renre studies. In A. Herrington & C. Moran (Eds.), *Genre across the curriculum* (pp. 1-18). Logan, Utah: Utah State University Press.

Hewings, M., & Hewings, A. (2002). "It is interesting to note that...": a comparative study of anticipatory `it' in student and published writing. *English for Specific Purposes, 21*(4), 367-383.

Hirano, E. (2009). Research article introductions in English for specific purposes: A comparison between Brazilian Portuguese and English. *English for Specific Purposes, 28*(4), 240-250.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes, 16*(4), 321-337.

Holtz, M. (2011). *Lexico-grammatical properties of abstracts and research articles: A corpus-based study of scientic discourse from multiple disciplines.* (Unpublished doctoral dissertation), Technische Universität Darmstadt, Darmstadt, Germany.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes, 7*(2), 113-121.

Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English- and Chinese-medium journals. *Journal of Pragmatics, 43*(11), 2795-2809.

Huang, P., & He, L. (2011). Comparative genre analysis of research methods in English and Chinese experimental research articles. *中国ESP研究, 2*(1), 22-27.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hüttner, J., Smit, U., & Mehlmauer-Larcher, B. (2009). ESP teacher education at the interface of theory and practice: Introducing a model of mediated corpus-based genre analysis. *System, 37*(1), 99-109.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics, 17*(4), 433-454.

Hyland, K. (1998a). *Hedging in scientific research articles*. Amsterdam: John Benjamins.

Hyland, K. (1998b). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics, 30*(4), 437-455.

Hyland, K. (1999). Talking to students: Metadiscourse in introductory coursebooks. *English for Specific Purposes, 18*(1), 3-26.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.

Hyland, K. (2002a). Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics, 34*(8), 1091-1112.

Hyland, K. (2002b). Genre, language, context, and literacy. *Annual Review of Applied Linguistics, 22*, 133-135.

Hyland, K. (2004). *Genre and second language writing*. Ann Arbor, MI: The University of Michigan Press.

Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173-192.

Hyland, K. (2008a). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K. (2008c). Disciplinary voices: Interactions in research writing. *English Text Construction, 1*(1), 5-22.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156-177.

Hyland, K., & Tse, P. (2005). Evaluative *that* constructions: Signaling stance in research abstracts. *Functions of Language, 12*(1), 39-63.

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly, 30*, 693-722.

IBM. (2008). Statistical Package for the Social Sciences (Version 16.0.2) [Computer Software].

Jacoby, S. (1986). *The inevitable forked tongue: An analysis of references to other researchers in literary research articles.* (M.A. Thesis), University of Birmingham, UK.

Johns, A. M. (2002). *Genre in the classroom: Multiple perspectives*. Mahwah, NJ: Lawrence Erlbaum.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kaltenböck, G., & Mehlmauer-Larcher, B. (2005). Coping with texts in the world of business: what a genre-informed approach can offer. *IATEFL/BESIG Business Issue, 3*(05), 9-12.

Kanoksilapatham, B. (2003). *A corpus-based investigation of scientific research articles: Linking move analysis with multidimensional analysis.* (Unpublished doctoral dissertation), Georgetown University, Washington, DC.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes, 24*(3), 269-292.

Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor & T. A. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 73-119). Amsterdam: John Benjamins.

Kanoksilapatham, B. (2009). *Writing research article abstracts in engineering: Linguistic perspective*. Paper presented at the International Conference on Science, Technology and Innovation for Sustainable Well-Being (STISWB), Mahasarakham University, Thailand.

Kanoksilapatham, B. (2011). Civil engineering research article Introductions: Textual structure and linguistic characterization. *The Asian ESP Journal, 7*(2), 55-84.

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural

variation in research articles across three engineering sub-discipline corpora.

English for Specific Purposes, 37, 74-86.

Kerlinger, F. N. (1979). *Behavioral research: Aconceptual approach*. New York: Holt,

Rinehart & Winston.

Kpolovie, P. (2011). Hypothesis postulation: the what, why, how and which. *Journal of

Education in Developing Areas, 19*(1), 1-15.

Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific

journal articles. *English for Specific Purposes, 18*(2), 121-138.

Kwan, B. S. C. (1996). *Introductions in state-of-the-art, argumentative, and teaching

tips TESL journal articles: Three possible sub-genres of introduction? Research

Monograph No 12*. Hong Kong: City University of Hong Kong.

Kwan, B. S. C. (2006). The schematic structure of literature reviews in doctoral theses

of applied linguistics. *English for Specific Purposes, 25*(1), 30-55.

Lave, J., & Wenger, E. (1991). *Situated learning: legitimate peripheral participation*.

Cambridge, UK: Cambridge University Press.

Lee, D. Y. W., & Chen, S. X. (2009). Making a bigger deal of the smaller words:

Function words and other key items in research writing by Chinese learners.

Journal of Second Language Writing, 18, 281-296.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistic, Proceedings of Nobel Symposium 82* (pp. 105-122). ACM, Berlin, New York: Mouton de Gruyter.

Leech, G. (1997). Introducing corpus annotation. In R. Garside, G. Leech & A. McEnery (Eds.), *Corpus annotation* (pp. 1-18). London: Longman.

Lewin, B. A. (2005). Hedging: An exploratory study of authors' and readers' identification of `toning down' in scientific texts. *Journal of English for Academic Purposes, 4*(2), 163-178.

Li, L.-J., & Ge, G.-C. (2009). Genre analysis: Structural and linguistic evolution of the English-medium medical research article (1985-2004). *English for Specific Purposes, 28*(2), 93-104.

Lim, J. M.-H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes, 25*(3), 282-309.

Lim, J. M.-H. (2010). Commenting on research results in applied linguistics and education: A comparative genre-based investigation. *Journal of English for Academic Purposes, 9*(4), 280-294.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Loi, C. K. (2010). Research article introductions in Chinese and English: A comparative genre-based study. *Journal of English for Academic Purposes, 9*(4), 267-279.

Loi, C. K., & Evans, M. S. (2010). Cultural differences in the organization of research article introductions from the field of educational psychology: English and Chinese. *Journal of Pragmatics, 42*(10), 2814-2825.

Lopez, G. (1982). *Article introductions in Spanish: A study in comparative rhetoric.* (M.A. Thesis), University of Aston, Birmingham, UK.

Luzon Marco, M. J. (2000). Collocational frameworks in medical research papers: A genre-based study. *English for Specific Purposes, 19*(1), 63-86.

Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet, 358*(9280), 483-488.

Martín-Martín, P. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes, 22*(1), 25-43.

Martínez, I. A. (2008). Building consensus in science: Resources for intertextual dialog in biology research articles. *Journal of English for Academic Purposes, 7*(4), 268-276.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*(3), 183-198.

Martín-Martín, P. (2002). A genre-based investigation of abstract writing in English and Spanish. *Revista canaria de estudios ingleses*(44), 47-64.

Martin, J. R. (1985). Process and text: Two aspects of human semiosis. *Systemic perspectives on discourse, 1*, 248-274.

Martin, J. R. (1992). *English text: System and structure*. Amsterdam: John Benjamins.

McEnery, A., & Wilson, A. (2001). *Corpus linguistics*. Edingburg: Edingburg University Press.

Melander, B., Swales, J. M., & Fredrickson, K. M. (1997). Journal abstracts from three academic fields in the United States and Sweden: National or disciplinary proclivities? In A. Duszak (Ed.), *Culture and Styles of Academic Discourse* (pp. 251-272). Berlin: Mouton de Gruyter.

Mosteller, F., Nave, B., & Miech, E. J. (2004). Why we need a structured abstract in education research. *Educational Researcher, 33*(1), 29-34.

Mur-Dueñas, P. (2007a). 'I/we focus on…': A cross-cultural analysis of self-mentions in business management research articles. *Journal of English for Academic Purposes, 6*(2), 143-162.

Mur-Dueñas, P. (2007b). Same genre, same discipline; however, there are differences: a cross-cultural analysis of logical markers in academic writing. *ESP Across Cultures, 4*, 37-53.

Mur-Dueñas, P. (2010). Attitude markers in business management research articles: A cross-cultural corpus-driven approach. *International Journal of Applied Linguistics, 20*(1), 50-72.

Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics, 43*(12), 3068-3079.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for Specific Purposes, 16*(2), 119-138.

Ozturk, I. (2007). The textual organisation of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes, 26*(1), 25-38.

Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. *Applied Linguistics, 15*(3), 288-299.

Parkinson, J. (2011). The discussion section as argument: The language used to prove knowledge claims. *English for Specific Purposes, 30*(3), 164-175.

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System, 30*(4), 479-497.

Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Peng, J. (1987). Organizational features in chemical engineering research articles. In T. Dudley-Evans (Ed.), *Genre analysis and ESP (ELR Journal 1)* (pp. 79-116): The University of Birmingham.

Pho, P. D. (2008a). How can learning about the structure of research articles help international students? In T. McGrath (Ed.), *Conference Proceedings of the 19th ISANA International Education Conference, 2-5 December 2008, paper 14*.

Pho, P. D. (2008b). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies, 10*(2), 231-250.

Pho, P. D. (2009). An evaluation of three different approaches to the analysis of research article abstracts. *Monash University Linguistics Papers, 6*(2), 11-26.

Pho, P. D. (2010). Linguistic realizations of rhetorical structure: A corpus-based study of research article abstracts and introductions in applied linguistics and educational technology. In S. T. Gries, S. Wulff & M. Davies (Eds.), *Corpus-linguistic applications: Current studies, new directions* (pp. 135-152). Amsterdam: Editions Rodopi B.V.

Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes, 18*(2), 139-160.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Riley, K. (1991). Passive voice and rhetorical role in scientific writing. *Journal of Technical Writing and Communication, 21*(3), 239-257.

Rodman, L. (1991). Anticipatory it in scientific discourse. *Journal of Technical Writing and Communication, 21*(1), 17-27.

Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes, 11*(2), 93-113.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*(2), 149-170.

Samraj, B. (2002). Introductions in research articles: variations across disciplines. *English for Specific Purposes, 21*(1), 1-17.

Santini, M., Mehler, A., & Sharoff, S. (2010). Riding the rough waves of genre on the Web: Concepts and research questions. In A. Mehler, S. Sharoff & M. Santini (Eds.), *Genres on the Web: Computational models and empirical studies* (pp. 3-30). Dordrecht: Springer.

Scott, D., & Morrison, M. (2005). *Key ideas in educational research*. New York, NY: Continuum.

Shehzad, W. (2007). How to end an introduction in a computer science article? A corpus-based approach. In E. Fitzpatrick (Ed.), *Corpus linguistics beyond the word: Corpus research from phrase to discourse* (pp. 227-241). Amsterdam: Editions Rodopi B.V.

Sheldon, E. (2011). Rhetorical differences in RA introductions written by English L1 and L2 and Castilian Spanish L1 writers. *Journal of English for Academic Purposes, 10*(4), 238-251.

Sinclair, J. (1987). Collocation: A progress report. In R. J. Steele & T. Threadgold (Eds.), *Language topics: essays in honour of Michael Halliday*. Amsterdam: John Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Stoller, F. L., & Robinson, M. S. (2013). Chemistry journal articles: An interdisciplinary approach to move analysis with pedagogical aims. *English for Specific Purposes, 32*(1), 45-57.

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes, 2*(4), 327-341.

Stubbs, M. (1995). Collocations and semantic profiles: on the cause of the trouble with quantitative methods. *Functions of Language, 2*, 1-33.

Stubbs, M. (2007). Quantitative data on multi-word sequences in English: the case of the word world. *Text, Discourse and Corpora: Theory and Practice*, 163-189.

Swales, J. M. (1981). Aspects of article introductions. Birmingham, UK: The University of Aston, Language Studies Unit.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research Genres: Exploration and applications*. Cambridge, UK: Cambridge University Press.

Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.

Swales, J. M., & Najjar, H. (1987). The writing of research article introductions. *Written Communication, 4*(2), 175-191.

Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1981). On the use of the passive in two astrophysics journal papers. *The ESP Journal, 1*(2), 123-140.

Tessuto, G. (2015). Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over. *English for Specific Purposes, 37*, 13-26.

Thomas, S., & Hawes, T. P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes, 13*(2), 129-148.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.

Thompson, D. K. (1993). Arguing for experimental "facts" in science: A study of research article results sections in biochemistry. *Written Communication, 10*(1), 106-128.

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*(4), 365-382.

Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for Academic Purposes. *Language Learning & Technology, 5*(3), 91-105.

Thomson Reuters. (2009). Top journals in education: Ranked by 2007 impact factor. Retrieved June 30, 2011, from http://sciencewatch.com/dr/sci/09/mar15-09_ 1D/

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins Publishing Co.

Toutanova, K., Klein, D., Manning, C., Morgan, W., Rafferty, A., Galley, M., & Bauer, J. (2015). Stanford POS Tagger (Version v3.5.2). Boston. Retrieved from http://nlp.stanford.edu/software/tagger.shtml

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). *Feature-rich part-of-speech tagging with a cyclic dependency network.* Paper presented at the HLT-NAACL 2003, Edmonton, Canada.

Tribble, C. (1997). *Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching*. Paper presented at the PALC 1997 Practical applications in language corpora, Lodz. http://www.ctribble.co.uk/text/Palc. Htm

Tribble, C. (2001). Small corpora and teaching writing: Towards a corpus-informed pedagogy of writing. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 381-408). Amsterdam: John Benjamins.

Trimble, M., & Trimble, L. (1982). Rhetorical-grammatical features of scientific and technical texts as a major factor in written ESP communication. In J. En Hoedt, L. Lundquist, H. Picht & J. Qvistgaard (Eds.), *Proceedings of the third European symposium on LSP, Copenhagen 1981: Pragmatics and LSP* (pp. 199-216). Copenhagen: Villadsen & Christensen.

United States Department of Education. (2015). Educational resources information center. Retrieved August 10, 2015 http://eric.ed.gov/?journals

Venuti, L. (2009). Translation, intertextuality, interpretation. *Romance Studies, 27*(3), 157-153.

Weissberg, R., & Buker, S. (1990). *Writing up research*. Englewood Cliffs, N.J.: Prentice Hall Regents.

Widdowson, H. G. (1998). Context, community, and authentic language. *TESOL Quarterly, 32*(4), 705-716.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 3-25.

Williams, G. C. (1998). Collcational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Applied Linguistics, 3*, 151-170.

Williams, I. A. (1996). A contextual study of lexical verbs in two types of medical research report: Clinical and experimental. *English for Specific Purposes, 15*(3), 175-197.

Williams, I. A. (1999). Results sections of medical research articles: Analysis of rhetorical categories for pedagogical purposes. *English for Specific Purposes, 18*(4), 347-366.

Williams, I. A. (2005). Contrasting discourse styles in Spanish and English research article discussions: Evidence for progressive and retrogressive patterns. http://www.metmeetings.org/pagines/Abstracts/williamssat%

Wingard, P. (1981). Some verb forms and functions in six medical texts. In L. Selinker, E. Tarone & V. Hanzali (Eds.), *English for Academic and Technical Purposes: Studies in Honor of Louis Trimble* (pp. 53-64). Rowley, MA: Newbury House.

Wood, A. (1982). An examination of the rhetorical structures of authentic chemistry texts. *Applied Linguistics, 3*(2), 121-143.

Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes, 28*(4), 421-450.

Xiao, R., & McEnery, A. (2005). Two approaches to genre analysis. *Journal of English Linguistics, 33*(1), 62-82.

Xu, J., & Jia, Y. (2011). BFSU qualitative coder (Version 1.1) [Computer Software]. Bejing: National Research Center for Foreign Language Education, Beijing Foreign Studies University.

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes, 22*(4), 365-385.

Zhang, B., Thuc, Q. B. T., & Pramoolsook, I. (2012). Moves and linguistic realizations: English RA abstracts by Vietnamese agricultural researchers. *The Asian ESP Journal, 8*(3), 126-149.

Zhang, Y., & Hu, J. (2010). A genre-based study of medical research article introductions: A contrastive analysis between Chinese and English. *The Asian ESP Journal, 6*(1), 72-96.
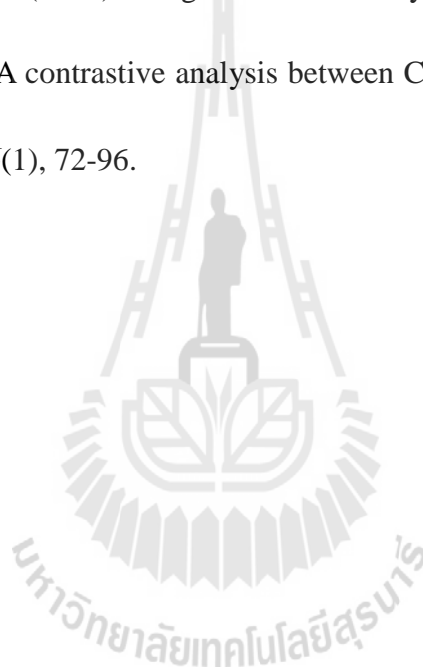
APPENDICES
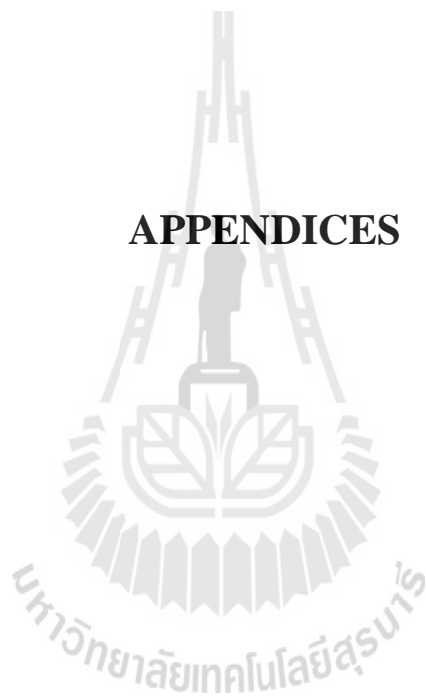
APPENDIX A

LIST OF JOURNALS FOR ERC

No.	Journal Title	Code	IF	Rank
1	Journal of Engineering Education	JEE	3.000	1
2	Academy of Management: Learning & Education	LAE	2.796	2
3	Scientific Studies of Reading	SSR	2.676	3
4	American Educational Research Journal	AER	1.930	6
5	Health Education Research	HER	1.786	7
6	Computers & Education	CAE	1.602	8
7	Journal of the Learning Sciences	JLS	1.571	9
8	Sociology of Education	SOE	1.290	13
9	Language Learning & Technology	LLT	1.222	14
10	Journal of Research in Science Teaching	RST	1.148	15
11	Elementary School Journal	ESJ	1.062	18
12	Learning and Instruction	LAI	1.029	19

APPENDIX B

SAMPLE OF MOVE-MARKED TEXT

(Excerpted from LAI_01)

<Title>Effects of different types of learning on handwriting movements in young children</Title>

<STtle>1. Introduction</STtle>

<IET2>Handwriting development has been investigated by several researchers who have demonstrated its close relationships with many different aspects of the child's development and environment (Chartrel & Vinter 2004; Graham, 2006; Karlsdottir, 1996; Rosenblum, Weiss, & Parush, 2004; Thomassen & Teulings, 1985; Van Galen, 1991). </IET2><IET4>Handwriting can be conceptualized as a perceptual-motor skill in which the perceptual component pertains to the letter shape and the motor component to the movement producing the letter trajectory.</IET4> <IET2>As the goal is to enable children to acquire fast and legible handwriting, a failure in this learning process is often associated, if not inevitably, to poor school performance. </IET2><IET3>Fayol and Miret (2005) have shown, for example, that children with poorer graphic skills performed worse in a dictation test. Thus, failure in learning to write letters involves negative consequences at the highest level of cognitive processes involved in text production (Olive, Favart, Beauvais, & Beauvais, 2009). Yet, all the studies estimate that approximately 10 – 20% of primary school children encounter difficulties learning to write (Alston, 1985; Maeland, 1992; Rubin & Henderson, 1982).</IET3>

APPENDIX C

PENN TREEBANK II TAG SET

TAG	DESCRIPTION	EXAMPLES
CC	conjunction, coordinating	*and, or, but*
CD	cardinal number	*five, three, 13%*
DT	determiner	*the, a, these*
EX	existential there	*there were six boys*
FW	foreign word	*mais*
IN	conjunction, subordinating or	*of, on, before, unless*
JJ	adjective	*nice, easy*
JJR	adjective, comparative	*nicer, easier*
JJS	adjective, superlative	*nicest, easiest*
LS	list item marker	
MD	verb, modal auxillary	*may, should*
NN	noun, singular or mass	*tiger, chair, laughter*
NNS	noun, plural	*tigers, chairs, insects*
NNP	noun, proper singular	*Germany, God, Alice*
NNPS	noun, proper plural	*we met two*
PDT	predeterminer	*both his children*
POS	possessive ending	*'s*
PRP	pronoun, personal	*me, you, it*
PRP$	pronoun, possessive	*my, your, our*
RB	adverb	*extremely, loudly,*
RBR	adverb, comparative	*better*
RBS	adverb, superlative	*best*
RP	adverb, particle	*about, off, up*
SYM	symbol	*%*
TO	infinitival to	*what to do?*

UH	interjection	*oh, oops, gosh*
VB	verb, base form	*think*
VBZ	verb, 3rd person singular	*she thinks*
VBP	verb, non-3rd person singular	*I think*
VBD	verb, past tense	*they thought*
VBN	verb, past participle	*a sunken ship*
VBG	verb, gerund or present	*thinking is fun*
WDT	*wh*-determiner	*which, whatever,*
WP	*wh*-pronoun, personal	*what, who, whom*
WP\$	*wh*-pronoun, possessive	*whose, whosever*
WRB	*wh*-adverb	*where, when*
.	punctuation mark, sentence	*. ; ? **
,	punctuation mark, comma	*,*
:	punctuation mark, colon	*:*
(	contextual separator, left paren	*(*
)	contextual separator, right	*)*

APPENDIX D

SAMPLE OF POS-TAGGED TEXT

(IDM_HER_09_01)

This_DT process_NN evaluation_NN adapted_VBD the_DT conceptual_JJ framework_NN described_VBN by_IN Linnan_NNP and_CC Steckler_NNP ._. Following_VBG an_DT extensive_JJ review_NN of_IN process_NN evaluation_NN research_NN ,_, they_PRP developed_VBD a_DT framework_NN for_IN systematically_RB conducting_VBG process_NN evaluations_NNS to_TO capture_VB key_JJ process_NN components_NNS :_: context_NN ,_, reach_NN ,_, dose_NN delivered_VBN ,_, dose_NN received_VBD and_CC fidelity_NN of_IN intervention_NN delivery_NN ._. `_`` Context_NN '_" assesses_VBZ aspects_NNS of_IN the_DT intervention_NN 's_POS environment_NN or_CC setting_VBG that_IN might_MD affect_VB its_PRP$ delivery_NN or_CC outcomes_NNS ._. `_`` Reach_VB '_" is_VBZ the_DT proportion_NN of_IN the_DT intended_JJ audience_NN who_WP actually_RB took_VBD part_NN in_IN the_DT intervention_NN ._. `_`` Dose_NN delivered_VBN '_" is_VBZ the_DT proportion_NN of_IN the_DT intended_JJ intervention_NN that_WDT was_VBD actually_RB provided_VBN to_TO the_DT target_NN audience_NN ._. `` Dose_NN received_VBD '_" is_VBZ the_DT extent_NN to_TO which_WDT the_DT intended_JJ audience_NN engaged_VBN with_IN and_CC used_VBD the_DT intervention_NN information_NN and_CC activities_NNS ._. `_`` Fidelity_NNP '_" is_VBZ the_DT quality_NN of_IN the_DT intervention_NN implementation_NN ,_, the_DT extent_NN to_TO which_WDT it_PRP was_VBD delivered_VBN in_IN the_DT intended_JJ manner_NN ._. Linnan_NNP and_CC Steckler_NNP additionally_RB discussed_VBD assessing_VBG recruitment_NN and_CC combining_VBG reach_NN ,_, dose_NN delivered_VBN ,_, dose_NN received_VBD and_CC fidelity_NN to_TO create_VB a_DT composite_JJ implementation_NN score_NN ._. We_PRP could_MD not_RB analyze_VB recruitment_NN as_IN a_DT key_JJ component_NN of_IN this_DT process_NN evaluation_NN because_IN we_PRP did_VBD not_RB collect_VB data_NNS on_IN reasons_NNS for_IN nonparticipation_NN ._. We_PRP also_RB did_VBD not_RB create_VB a_DT composite_JJ implementation_NN score_NN because_IN of_IN the_DT conceptual_JJ difficulty_NN of_IN determining_VBG the_DT relative_JJ weight_NN of_IN each_DT component_NN when_WRB combined_VBN ._.

APPENDIX E

INITIAL CODING SCHEME

INTRODUCTION	
Move 1	**Establishing a territory**
Step 1	Claiming centrality
Step 2	Making topic generalizations
Step3	Reviewing items of previous literature
Move 2	**Establishing a niche**
Step1A	Counter-claiming
Step1B	Indicating a gap
Step1C	Question-raising
Step1E	Continuing a tradition
Move 3	**Presenting the present work**
Step 1	Announcing the present research
Step 2	Announcing principal findings
Step 3	Indicating RA structure
METHODS	
Move 1	**Describing data collection procedures**
Step 1	Describing the sample
Step 2	Describing research instruments
Step 3	Elaborating on data collection procedures
Step 4	Justifying data collection procedures
Move 2	**Describing data analysis procedures**
RESULTS	
Move 1	**Preparing for the presentation of results**
Move 2	**Reporting specific results**
Move 3	**Commenting on specific results** (Interpreting results)
DISCUSSION	
Move 1	**Providing background information**
Move 2	**Highlighting overall research outcome**
Move 3	**Commenting on specific results**
Step 1	Interpreting results
Step 2	Comparing results with literature
Step 3	Accounting for results
Move 4	**Drawing conclusions of the study**
Move 5	**Evaluating the study**(Indicating limitations)
Move 6	**Deductions from the research**
Step 1	Drawing implications or making suggestions
Step 2	Recommending future research

APPENDIX F

LIST OF LINGUISTIC FEATURES

This is a list of the 67 linguistic features investigated in the study, with some features grouped into super-ordinate categories. Their operational definitions are found in (1988, pp. 221-245).

1. Past tense

2. Present perfect aspect

3. Present tense

4. Adverbs

    1)Total adverbs

    2) Place adverbials

    3) Time adverbials

5. Personal pronouns

    1) First person pronouns

    2) Second person pronouns

    3) Third person pronouns

6. Pronoun *it*

7. Demonstratives

    1) Demonstrative pronouns

    2) Demonstratives

8. Indefinite pronouns

9. Pro-verb *do*

10. Direct WH-questions
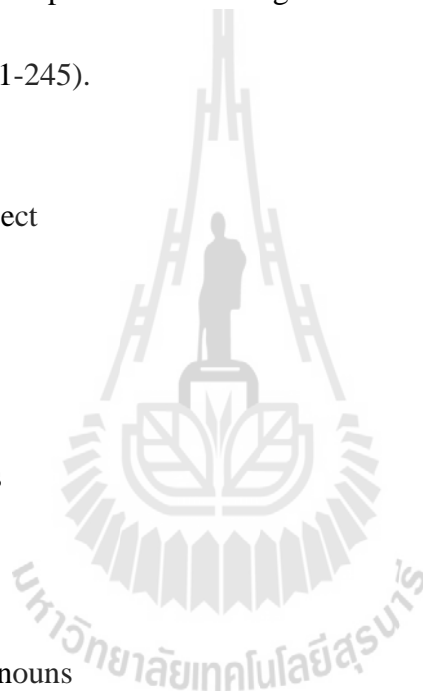
11. Nominalizations & gerunds

    1) Nominalizations

    2) Gerunds

12. Nouns

13. Passives

1)   Agentless passives

2)   By-passives

14. *Be* as main verb

15. Existential *there*

16. Complements

    1) *That* verb complements

    2) *That* adjective complements

    3) WH-clauses

17. *To*-infinitives

18. Participial clauses

1)   Present participial clauses

2)   Past participial clauses

3)   Past participial WHIZ deletion relatives

4)   Present participial WHIZ deletion relatives

19. Relative clauses

    1) *That* relative clauses on subject position

    2) *That* relative clauses on object position

    3) WH relative clauses on subject position

    4) WH relative clauses on object position

    5) Pied-piping relative clauses

    6) Sentence relatives

20. Adverbial clauses

1) Causative adverbial subordinators

2) Concessive adverbial subordinators

3) Conditional adverbial subordinators

4) Other adverbial subordinators

21. Prepositional phrases

22. Attributive adjectives

23. Predicative adjectives

24. Type-token ratio

25. Word length

26. Conjuncts

27. Pragmatic expressions

1) Downtoners

2) Hedges

3) Amplifiers

4) Emphatics

5) Discourse particles

28. Modals

1) Possibility modals

2) Necessity modals

3) Predictive modals

29. Public verbs

30. Private verbs

31. Suasive verbs

32. Seem/appear

33. Dispreferred forms

1) Contractions

2) Subordinator *that* deletion
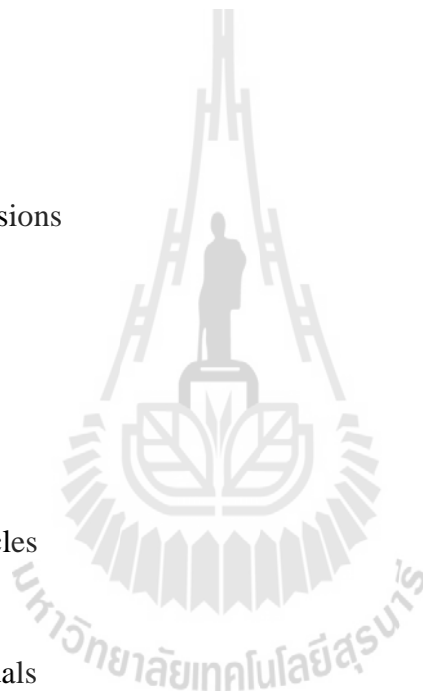
3) Stranded preposition

4) Split infinitives

5) Split auxiliaries

34. Phrasal coordination

35. Independent clause coordination

36. Negation

1) Synthetic negation

2) Analytic negation
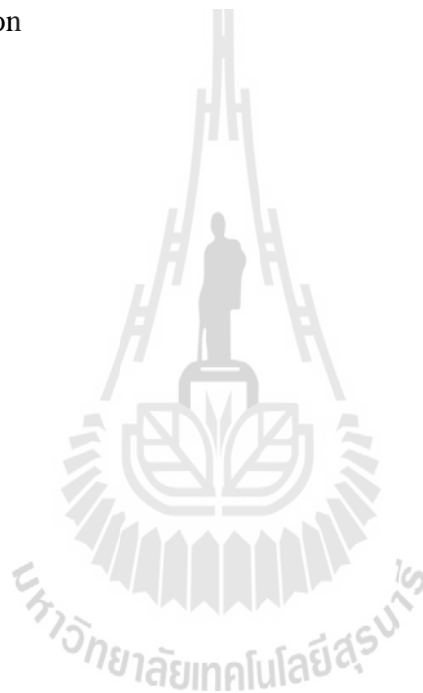
APPENDIX G

TOTAL VARIANCE EXPLAINED

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	3.346	9.294	9.294	3.346	9.294	9.294	2.424
2	2.429	6.748	16.042	2.429	6.748	16.042	2.699
3	1.838	5.105	21.146	1.838	5.105	21.146	2.329
4	1.493	4.148	25.295	1.493	4.148	25.295	1.994
5	1.434	3.984	29.279	1.434	3.984	29.279	2.051
6	1.337	3.715	32.994	1.337	3.715	32.994	1.870
7	1.311	3.643	36.637	1.311	3.643	36.637	1.670
8	1.218	3.382	40.020				
9	1.162	3.228	43.248				
10	1.125	3.124	46.372				
11	1.092	3.033	49.405				
12	1.088	3.023	52.428				
13	1.047	2.909	55.338				
14	1.021	2.836	58.174				
15	.989	2.746	60.920				
16	.970	2.694	63.615				
17	.953	2.646	66.261				
18	.936	2.600	68.861				
19	.908	2.523	71.383				
20	.880	2.444	73.828				
21	.873	2.426	76.254				
22	.854	2.373	78.627				
23	.798	2.218	80.845				
24	.787	2.186	83.031				
25	.766	2.127	85.158				
26	.757	2.104	87.261				
27	.706	1.960	89.221				
28	.690	1.916	91.137				
29	.635	1.763	92.900				
30	.608	1.689	94.589				
31	.537	1.490	96.080				
32	.514	1.427	97.506				
33	.411	1.142	98.649				
34	.235	.652	99.300				
35	.150	.415	99.715				
36	.102	.285	100.000				

Extraction Method: Principal Component Analysis

CURRICULUM VITAE

Mr. Baoya Zhang is currently an associate professor in the School of Foreign Languages, Chongqing Technology and Business University in Chongqing, China. He holds a BA degree in English Language and Literature from Central China Normal University in China, and a MA degree in Teaching Reading and Language Arts from Oakland University in the USA. After teaching high school English in Guizhou and Sichuan for ten years, he became a lecturer and later an associate professor at Guizhou Minzu University in Guizhou, China, where, over a period of seventeen years, he taught a wide range of courses including Advanced English, Academic Writing, English Linguistics, Translation, and Lexicology. He moved to the current job two years ago.

Mr. Zhang has published research in areas such as TEFL, children's literature, reading, and genre analysis. He also co-edited the award-winning "New Era Interactive English," a coursebook series for university students. His new research interests include multidimensional analysis and corpus linguistics.