MULTIPLE PERSPECTIVES ON WRITING FOR RESEARCH PUBLICATION ACROSS ACADEMIC FIELDS

BY

TAHA OMIDIAN

A thesis

submitted to Victoria University of Wellington in fulfilment of the requirements

for the degree of Doctor of Philosophy

School of Linguistics and Applied Language Studies

Victoria University of Wellington

2021

Abstract

Writing for research publication is one of the primary means of disseminating knowledge in contemporary academia. It is through this particular form of knowledge dissemination that a rich body of scientific knowledge about a given phenomenon is accumulated. The primary goal of this dissertation is to explore the linguistic characteristics of this specific sub-register of academic writing. For this purpose, the study adopts a multi-perspective approach through which patterns of language use in research writing are investigated at three different linguistic levels: lexis, multi-word expressions, and general lexico-grammatical items. In doing so, careful consideration are given to disciplinarity, intra-textual variation, and L1-LX expert writing as key parameters of variation in writing for research publication. A secondary goal of the dissertation is to explore and highlight the benefits of methodological triangulation in corpus linguistic research. To this end, linguistic patterns in this study are identified, triangulated and verified through various inductive corpus-based analytical techniques to offer multiple, complementary perspectives on the discourse of research writing across disciplines. The data analyses in this dissertation are based on a representative corpus of empirical research articles (c. 4.5 million words) from a wide range of disciplines: biology, chemistry, dentistry, physics, mechanical engineering, applied linguistics, business, management, politics, and sociology. The analyses rely on both quantitative and qualitative assessments to provide a better understanding of the identified patterns. The results of these analyses show that the language of research writing is characterized by highly specialized and conventionalized discourses that are far from being homogeneous. It is found that such characteristics have the potential to govern the delineation of authors' linguistic choices at various levels of language use such as lexis, multi-word expressions, and more complex lexico-grammatical linguistic structures. It is also found that any characterization of research writing with no careful consideration given to such factors as disciplinary writing and intra-textual variation would be

incomplete. The results of the analyses also reveal considerable differences between L1 and LX expert writing in fulfilling the communicative practices related to the evaluation and elaboration of research findings.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1. Aims of the thesis

Writing for research publication is one of the most central activities of academic institutions. Scholars in various academic fields share and disseminate the outcome of their scientific endeavor through writing and publishing academic journal articles. This particular form of knowledge dissemination in academia can often present great challenges for those who wish to participate in scholarly activities in their fields but have not yet acquired the discourse competencies required for such a feat. These challenges are mostly rooted in the lack of familiarity with the ways in which ideas and arguments are conventionally organized and presented in different academic disciplines. Research has emphasized that academic texts, even within a single genre such as the research article, can show strong variation across different academic disciplines (Hyland, 2008; Hyland & Tse, 2007; Durrant, 2017). Hyland (2013) argues that such differences highlight the heterogeneous nature of academic writing, and that gaining literacy in such a skill requires a strong understanding of the particularities that characterize different disciplinary discourses. Developing such an understanding, however, demands prolonged exposure to the conventions and norms of a given academic community (Hyland, 2013). The challenges of acquiring such disciplinary knowledge become even more compounded for second language writers who use English for research publication purposes (see Flowerdew, 2013). This particular group of language users needs to make additional effort to not only become acquainted with the intricacies of academic writing in a L2 but also to adhere to the linguistic expectations of the target readership in their fields. Investigating writing for research publication would, therefore, seem central not only for

gaining a nuanced understanding of how knowledge is typically constructed and negotiated across disciplines but also for addressing the challenges involved in using English for Specific Purposes.

Research into research writing has been guided by two main approaches. One approach has been to study the discursive and textual structure of academic genres across disciplines. This approach, which was spearheaded by Swales' (1990) seminal work on genre analysis, has been mainly focused on the rhetorical strategies that characterize disciplinary practices (Basturkmen, 2012; Bruce, 2008; Samraj, 2002; Stoller & Robinson, 2013). Although this line of research has offered important insights into the textual organization of academic genres and how they may differ across disciplines, it has been criticized for its unsystematic (and mainly qualitative) approach to studying academic texts (Cortes, 2015; Crawford & Csomay, 2015; Hyland, 2013; Paltridge, 1994). As Biber, Connor, and Upton (2007, p. 36) argue, due to the qualitative nature of this approach, analysts often resort to analyzing a small number of academic texts which, in turn, limits the generalizability of their findings.

A different approach altogether has been to study disciplinary variation through the analysis of the linguistic features that distinguish academic disciplines. This approach, which was made possible by advances in corpus methodologies in the late 1970s, focuses on linguistic variation in a large and principled collection of texts across various disciplines. Adopting this approach, a multitude of corpus-based studies have investigated disciplinary variation by looking at single words (Durrant, 2014; Jiang & Hyland, 2017) and multi-word expressions (Durrant, 2017; Hyland, 2008) in various academic genres. This body of research is also complemented by investigations of general patterns of lexico-grammatical features across disciplines and various academic genres, such as Biber's (1988) Multi-dimensional analysis.

Although the findings from these two lines of inquiry have greatly informed the theoretical and applied research into academic writing across disciplines, they have shared certain key constraints that have made it difficult to arrive at a comprehensive view of research writing. First, the majority of these studies have focused on only one section of this genre (e.g., Basturkmen, 2012; Cortes, 2013; Omidian, Shahriari, & Siyanova-Chanturia, 2018). Specifically, very few studies have considered a cross-sectional analysis of research articles and the impact of intratextuality on research writing across different disciplines (e.g., Parkinson, 2013; Samraj, 2005; Stoller & Robinson, 2013). This is surprising because studies have long emphasized the importance of considering possible intra-textual variations within academic genres (e.g., Bhatia, 1991; Swales, 2004; Swales & Feak, 2004). As Biber (1988, p. 171) argues, "academic prose texts can be quite different from one another and still be considered representative of their genre".

Secondly, studies of variation in research writing have largely overlooked the linguistic differences that might exist in the writing practices of L1 and LX[1] scholars across sections of research articles in different disciplinary fields. Highlighting such differences is important because the ways in which L1 and LX expert writers reflect the shared conventions of their academic fields might not necessarily be similar. Additionally, with increasing pressure on non-native scholars in different academic fields to publish their research in English, more scientific publications are now produced by LX authors. This has resulted in on-going changes in the norms of language use in academic writing (Mauranen, Hynninen & Ranta, 2016, p. 44). As Perez-Llantada (2014, p. 192) argues, LX writers often consciously (or unconsciously) merge their culture-specific linguistic features with Anglophone norms and craft new conventions in academic writing. Very few studies have attempted to take

---

[1] In the present study, we use the term 'LX' (proposed by Dewaele, 2018) to refer scholars whose first language is not English (for further discussions of this term and its advantages over other commonly used labels such as L2 and non-native speakers, see Dewaele, 2018).

account of these different (but not necessarily infelicitous) forms of language use in LX expert writing (Martinez, 2018; Pan, Reppen, Biber, 2016). However, these studies have largely overlooked the role of disciplinarity and intratextuality as important sources of variation in research writing.

Finally, almost all studies in this field have narrowed their focus to the analysis of either single or multi-word items (but not both). However, as is well known, vocabulary is composed of both single and multi-word items (Gardner, 2007; Nattinger & DeCarrico, 1992; Schmitt, 2014; Sinclair, 1991; Wray, 2002). In addition, it is possible that variation in research writing would manifest itself in various co-occurrence patterns of lexical and grammatical linguistic devices (e.g., that-clauses controlled by stance verbs). Therefore, it is not unreasonable to argue that adopting multiple linguistic perspectives would potentially result in a more comprehensive view of research writing across disciplinary fields. However, to my knowledge, no study to date has adopted such a multi-perspective approach to explore various linguistic characteristics of research writing across academic fields.

Acknowledging the above-mentioned gaps, the present research seeks to explore the language of research writing across academic disciplines by adopting a multi-perspective approach through which through which patterns of language use were investigated from three different linguistic perspectives: lexis, multi-word expressions, and lexico-grammatical items. For this purpose, three studies are designed, each adopting one of the above linguistic perspectives. Specifically, the first study focuses on exploring parameters of variation in the use of single-word vocabulary across different parts of research articles in different disciplinary fields. The second study investigates such variation in the correspondence between form and function of multi-word expressions. The third study explores the linguistic dimensions characterizing the co-occurrence patterns of lexico-grammatical features of

research writing. This study also aims to investigate possible variation in the writing practices of L1- and LX- English scholars in sections of research articles across disciplines.

1.2. Research questions

This proposed research seeks to investigate disciplinary influences on the use of linguistic features that characterize the communicative aims of different sections of the RA (i.e., Introduction, Methods, Results, Discussion, and Conclusion). It will look at such disciplinary characteristics through the lens of single words, multi-word items, and co-occurrence patterns of lexical and grammatical linguistic devices. This research also aims to unravel how L1- and LX-English authors may differ in their use of such linguistic features. Specifically, the following research questions will be addressed:

i. Are there systematic differences in the use of single word items used by authors in different academic fields across different sections of the RA?

ii. To what extent do the form-function mappings of multiword expressions frequently used to perform communicative aims of different sections of research articles vary as a result of disciplinary influences?

iii. To what extent do academic disciplines influence the co-occurrence patterns of lexical and grammatical linguistic devices (e.g., *that* verb complements, possibility modals, agentless passive) in different sections of the RA?

iv. Does native speaker status have a significant influence on the writing practices of L1- and LX-English scholars, as it is reflected in the co-occurrence patterns of lexical and grammatical devices commonly used in sections of research articles?

1.3. Outline of the dissertation

The present dissertation comprises seven chapters. Chapter 2 discusses key notions pertaining to disciplinarity and research writing and explains the inextricable entanglement of disciplinary knowledge and research writing competence. Chapter 3 explains methodological procedures undertaken to build the corpus used for the purposes of this study. Chapters 4, 5, and 6 report three separate corpus-based studies, describing the linguistic analyses conducted in each and discussing the findings that emerged from these quantitative and qualitative analyses. More specifically, Chapter 4 discusses the aims, methods, and results of a study (Study 1) focused on exploring the parameters of variation in the language of research writing as reflected in the use of single-word vocabulary. The study adopts an inductive approach through which patterns of specificity in vocabulary use are systematically identified and verified using different methods of data analysis. Chapter 5 discusses the results of a cross-disciplinary investigation of variation in the form-function mappings of multi-word expressions in research writing. The overarching aim of the study (Study 2) is to investigate the connection between form and function in research writing across disciplines. The study employs a two-stage mixed-methods research design in which linguistic patterns are first quantitatively analyzed and then further examined by a qualitative appraisal. Chapter 6 reports the results of a multi-dimensional analysis of the corpus. The study (Study 3) seeks to provide a multi-dimensional model of research writing across disciplinary fields. In doing so, the study explores and uncovers the principal linguistic dimensions associated with co-occurrence patterns of lexico-grammatical features of research writing. The study also aims to unravel how L1- and LX-English authors may differ in their use of such linguistic features. Finally, Chapter 7 summarizes and discusses the findings that emerged from the corpus-based analytical methods used in conducting the three studies. In doing so, the chapter highlights the advantages of the multi-perspective research approach adopted in this study and discusses

the benefits of such methodological triangulation for corpus-based research in general. The chapter also discusses the limitations of the research and concludes with a discussion of its implications for understanding the distinct linguistic characteristics and the discourse conventions of writing for research publication across academic fields.

Chapter 2

Literature review

2.1. Overview of the chapter

The chapter opens with an account of how the institutionalization of formal knowledge in the late Victorian era gave rise to the modern system of academic disciplines. It then goes on to explain how the intellectual boundaries of disciplinary communities can shape the ways in which knowledge is constructed and negotiated through the medium of written language in academia. The chapter concludes with an argument concerning the entanglement of disciplinary knowledge and research writing competence.

2.2. The genesis of disciplinarity

The late Victorian era has long been associated with developments in the institutionalization of the human sciences (Anderson & Velente, 2002, Bender, 1997; Russell, 1991, Shumway & Messer-Davidow, 1991). During this formative period, many educational systems (especially in Europe) saw fundamental changes in their social structures and institutional practices. Anderson and Velente (2002), in an edited collection of historical essays entitled *Disciplinarity at Fin de Siècle,* highlight the important role of the educational reform practices of this era in the emergence of discrete intellectual fields in academe. The majority of these practices were mainly related to the compartmentalization of 'scholarly knowledge', as the main product of formal education and academia at large (e.g. see Hoskin, 1993; Russell, 1991). The philosophy which drove these changes was motivated by the recognition that scholarly communities in academia are so diverse and disparate that providing an overarching, unified explanation for the different forms of knowledge they produce is

virtually impossible (Butcher & Strauss, 1961; also see Mannheim, 1936, p. 67 for a detailed discussion of the sociology of knowledge). Included in this recognition was the view that the creation of knowledge is highly likely to be differentially affected by the epistemological orientations of communities and their routine scholarly practices. And thus, accomplishing the task of structuring and organizing such a multi-dimensional phenomenon as knowledge appears to be symbiotically linked with the ability to govern the delineation of academic communities and characterize their intellectual practices.

As a result of this recognition, the closing decades of the nineteenth century saw numerous sociologists and philosophers of knowledge attempting to provide detailed descriptions of academic communities and their distinguishing characteristics (see Anderson & Velente, 2002). In essence, scholarly communities are typically viewed as social systems, operating on conventionally established codes that are shared by members (Clark, 1983; Hyland, 2015). These social and professional codes represent accepted values, research and discourse practices, and theoretical orientations of any given community of scholarship. Scholarly communities, in other words, provide a shared professional context within which specialized knowledge is produced by adhering to community-valued practices (Hyland, 2015). They are the places where members (as potential producers of knowledge) explore, debate, clarify and resolve technical and professional issues to push the boundaries of human understanding of different phenomena.

In theory, such conceptualizations could be used to define and differentiate the scholarly activities of communities in academia. However, in practice, the boundaries drawn by these initial conceptualizations have been found to be fuzzy (e.g., see Becher, 1994). Marking the boundaries of communities has always been a contested issue and has occasioned longstanding debates and disputes among sociologists (see Clark, 1983). This is perhaps due to the fact that knowledge communities are not often entirely fixed or

harmonious. These communities are constantly adjusting to recent trends in their field of knowledge and responding to the demands placed on them by exogenous forces such as universities and society as a whole (Post, 2009). This means that communities in academia are continually changing, evolving, and diversifying. This natural tendency of scholarly communities could, in turn, complicate the task of defining their characteristics. As Becher and Trowler (2001) maintain, the complex and ever-changing nature of academic communities means that there is no single method of enquiry or standard verification procedure that could uniquely characterize these social entities (p. 65).

Gradually, however, it has become clear that what distinguishes scholarly activities from one another is the manner in which communities conceptually define the content of their domains, the ways in which they formulate their research plans, and the principles of discovery and verification they use to create and test knowledge (Shulman, 1981, p. 6). Thus, when we speak of community-specific knowledge, we speak not merely of a body of information, but a set of community-specific theories, procedures, methods, and strategies, based on which knowledge is produced. It was this nuanced understanding of scholarly endeavor that led to the "reordering of intellectual life – the creation of academic disciplines" in the last third of the nineteenth century (Bender, 1997, p. 34). This reordering of knowledge resulted in the breakup of natural and moral philosophy into independent natural and social disciplines, respectively (Shumway & Messer-Davidow, 1991). In the early twentieth century, those disciplines that were sufficiently separate from the natural and social sciences were grouped together to form "Humanities" (ibid.). The emergence of disciplinarity also contributed to the creation of departments, as the principal bases of knowledge production within universities (Straus, 1973). Institutions comprising multiple segments (i.e., departments), each operating on well-organized, accepted body of knowledge and practices could produce knowledge more efficiently than those struggling with organizing conflicting

perspectives and ambiguous materials (Clark, 1987). This is the reason why most university departments are often the renamed versions of the disciplines they represent (e.g., the department of economics and finance, the department of biological sciences). According to Clark (1983, p. 30), academic disciplines are also the dominant force in the working lives of academics. As Clark (1983) argues, the primacy of disciplines in academia is so pronounced that if an academic worker is given the choice of leaving the discipline or institution, he or she will typically opt for leaving the institution.

Altogether, disciplines can be viewed as a specialized form of organization that knit knowledge, its producer (e.g., scholars), and the place of its production (e.g., institutions) together. Disciplines, therefore, mark a crucial point in the institutionalization of scholarly knowledge and the practices involved in creating it in academia (Fuller, 1991).

2.3. Disciplinary specificity

2.3.1. Typology of disciplinary knowledge

Disciplines are said to provide academic institutions with a contextualized framework that guides their research activities (Clark, 1983). In theory, the intellectual boundaries drawn by disciplinary frameworks allow members to pursue various academic goals within the same clearly defined context. However, in practice, these boundaries are not always straightforward or easy to follow. This is partly due to the epistemological overlaps that exist between certain academic areas. As Becher (1994) argues, disciplines are distinguished by the shared features of their associated domains of enquiry and, in very many instances (e.g., Biology and Chemistry or Linguistics and Psycholinguistics), these domains tend to overlap with those of neighboring fields. Many studies have attempted to account for such similarities by proposing different classification schemes for academic areas (Becher, 1989; Biglan, 1973a; Kolb 1981; Kuhn, 1970; Pantin, 1968).

In one of the earliest of such studies, Pantin (1968) explored the knowledge structures of the natural sciences. Based on impressionistic evidence gathered through investigating various phenomena of interest across the natural sciences, Pantin (1968) marked the contrasts between such disciplines by classifying them into *restricted* and *unrestricted* sciences. In Pantin's views, there is a real, and graded, distinction between disciplines such as the biologies (e.g., Geology, Zoology) and the physical sciences (e.g., Physics, Mechanical Engineering). The former, as he elaborates, are unrestricted sciences in the sense that their investigators are prepared to follow their research issues into any other sciences. The latter, in contrast, are mainly restricted to the field of knowledge to which they are devoted, such that the investigator in these fields is not often required to travers all other sciences. Pantin goes on to argue that the less-divergent fields of enquiry in the restricted sciences entail the use of certain standardized methods which are often unique and not typical of all the sciences. Included in this reasoning is the recognition that academic disciplines focusing on a more restricted and less diverse field of phenomena often exhibit specific characteristics which set them apart from other fields.

In a somewhat similar study, Kuhn (1970) explored the structure of revolutionary phases in the sciences. Kuhn's survey of the history of science led him to develop the concept of *paradigm.* By paradigm, Kuhn refers to a set of recurrent and quasi-standard theories that guide the conceptual, observational, and instrumental applications of research in a given field. As he writes:

> The study of paradigm is what mainly prepares the student for membership in the particular scientific community with which he will later practice. Because he there joins men who learned the bases of their field from the same concrete models, his subsequent practice will seldom evoke overt disagreement over fundamentals. (Kuhn, 1970, p. 10)

He further elaborates on the point:

> Men whose research is based on shared paradigms are committed to the same rules
> and standards for scientific practice. That commitment and the apparent consensus it
> produces are prerequisites for normal science, i.e., for the genesis and continuation of
> a particular research tradition. (ibid)

To Kuhn, the existence of a clearly established paradigm and the unique and esoteric type of research it permits can be taken as a sign of maturity in the development of any given scientific field. This particular view of paradigm in academic fields plays an important role in how Kuhn's differentiates disciplinary groupings. In Kuhn's view, disciplinary fields can be classified into two main categories: pre-paradigmatic and paradigmatic. As Kuhn explains, the former group of disciplines is often on the earlier, or pre-paradigm, stages of development. Pre-paradigmatic disciplinary communities are typically marked by high level of internal disagreements. Here members are drawn from a wide range of knowledge bases, each attempting to establish a new paradigm in their field, which in turn leads to diversity and dissension within the community. In contrast, disciplines with clear and unambiguous paradigms (i.e., paradigmatic disciplines) often exhibit high degree of disciplinary consensus and solidarity, as members working in such fields operate within an already established framework of understanding. Kuhn (1970) does not discuss specific examples of preparadigmatic and paradigmatic disciplines, but he designates physical and biological sciences as research areas that are guided by clearly established paradigms. Kuhn's account of disciplinary specificity is important as it provides a clear indication of how certain established practices in disciplinary communities can be used as demarcation criteria to draw definable boundaries between academic disciplines.

As is evident from the above, the taxonomies proposed by Pantin and Kuhn share certain similarities. First, both are concerned primarily with the sciences. Second, both

taxonomies involve polar opposites (restricted vs. unrestricted, preparadigmatic vs. paradigmatic). It should also be noted that both Pantin and Kuhn based their categorizations on impressionistic comparisons drawn by observing the variation in how research is typically conducted across disciplinary fields.

Adopting a more empirical approach to characterizing disciplinary fields in academia, Biglan (1973a) analyzed the judgments of faculty members on the shared characteristics of different academic subject areas at two universities in the United States. One hundred sixty-eight faculty members at the University of Illinois made judgments about the similarities of the subject matter of 36 areas, and 54 scholars at a college judged similarities among 30 areas. The judges were required to put academic areas into categories on the basis of their similarities. Using this method of sorting, the judgments of each faculty member about the similarities among subject areas were represented in a symmetrical matrix of overlap. Following this step, Biglan (1973a) performed multidimensional scaling of subject matter characteristics on the basis of the matrices created for scholars in each of the two universities. The emerged 'dimensions' were then correlated with ratings of the attributes of each area. The results of these analyses revealed three dimensions to be involved in shaping academics' perceptions of disciplines. These dimensions included (a) the existence of an established body of theory, or lack thereof, in certain disciplines (hard vs soft), (b) the degree to which the area in question has applications for practical and real-world issues (pure vs. applied), and (c) the extent to which it deals with inanimate objects (life system vs. non-life system).

Using these three dimensions, Biglan (1973a) grouped the disciplines in his study under three main academic clusters: hard-soft, applied-pure, and life-nonlife systems, respectively. It is important to note that the most prominent dimension (in terms of the variability it accounted for) in Biglan's study was the hard-soft cluster, which clearly distinguished subject matters such as engineering and agriculture from social sciences,

education, and humanities. Biglan (1973a) found that the key contrast between the hard sciences and soft disciplines lies in the existence a paradigm (or lack thereof) in these fields. According to Biglan (1973a), paradigms serve an important organizational function by providing a consistent account of different phenomena of interests (and their corresponding research issues) in an academic field. By studying and engaging with these field-specific paradigms, which are often discussed in textbooks, lectures, and laboratory exercises, the members of the corresponding scholarly community can formulate and prepare their research endeavor according to the established and developing body of knowledge in their field (Kuhn, 1970). Biglan's (1973a) findings revealed that the prominence of such paradigms in natural sciences was the differentiating factor between the hard and soft fields investigated in his study. He also found that paradigmatic fields (e.g., Engineering, Physics, Biology) were often characterized by greater consensus about appropriate content and method than were nonparadigmatic fields (e.g., Education, Linguistics, Politics). Biglan's findings indicate that the boundaries of knowledge are less clearly defined and are fuzzier in soft than hard knowledge fields. As will be discussed later, the fuzzy boundaries of knowledge and lack of a clear research paradigm in humanities and social sciences can profoundly impact the ways in which writers in these fields present their research endeavor.

Where Biglan (1973a) based his disciplinary classification on faculty members' judgements and perceptions of knowledge fields and their unique characteristics, Kolb (1976) used students' learning strategies as the basis for classifying disciplinary areas in academia. Using a psychometric test (known as Learning Style Inventory), Kolb (1976) set out to investigate individual learning styles along two primary dimensions: abstract-concrete and active-reflective. The former represents a continuum, with the ability to integrate observations into sound theories (abstract conceptualization) at one extreme and the tendency to fully engage in new experiences and immerse in one's immediate experience (the concrete

experiencing of events) on the other. The latter dimension (active-reflective) represents active experimentation (using theories to solve problems and actively testing the implications of one's hypothesis) at one end and reflective observation (relying on detached observations and reflectively interpreting data already collected) at the other. Focusing on these two learning dimensions (abstract-concrete vs active-reflective), Kolb (1976) examined learning styles of 800 practicing managers and graduate students in Management, who came from across a wide variety of disciplinary and educational backgrounds. Kolb found that, although participants shared a common occupation/profession, their leaning style and strategies considerably varied were according to their undergraduate educational experience. For example, it was found that students coming from engineering backgrounds had abstract-active learning styles, whereas political science, history and English majors had concrete-reflective learning styles. Kolb's (1976) findings indicated that disciplinary background can be considered a differentiating factor in shaping individual learning styles, "whether by the process of selection into a discipline, or by socialization in the course of learning in that disciplined, or both" (Kolb, 1981, p. 239).

In addition, Kolb's disciplinary dimensions overlapped considerably with those found in Biglan (1973a). Focusing on Biglan's hard-soft and pure-applied dimensions, Kolb (1981) showed that the majority of the disciplines that were common to the two studies were in identical categories. The substantial overlap between Biglan's disciplinary classification and the clustering of data on Kolb's (1976) learning dimensions (i.e., abstract-concrete and active-reflective) is shown in figure 2.1.
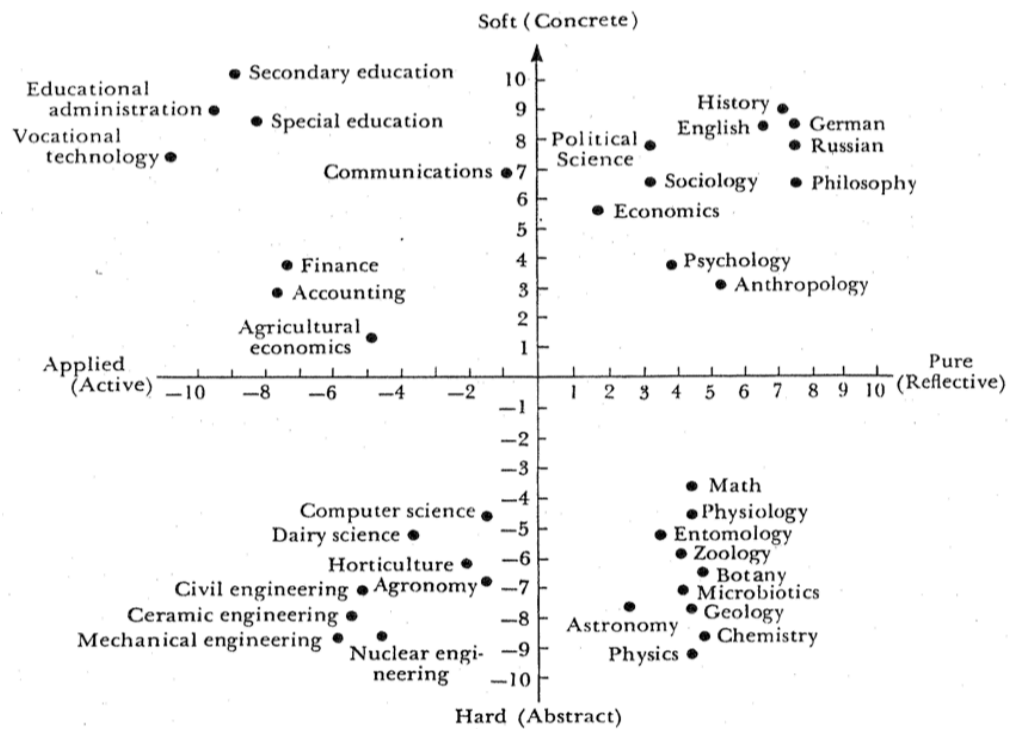
Figure 2.1. Biglan-Kolb disciplinary classification (adopted from Kolb, 1981)

Kolb (1981) replicated the results of Kolb (1976b) using a much larger data set. Based

on questionnaire data collected from graduate students as well as faculty members, Kolb

(1981) created ad hoc indices of the two learning dimensions (i.e., abstract-concrete and

active-reflective) for 45 academic fields. The abstract-concrete index was calculated based on

the percentage of graduate students who strongly agreed that educational experience in either

mathematics or humanities was important for their fields. The active-reflective index was

computed based on the percentage of faculty members who often provided paid consultation

to government, business, or other organizations – this was taken as an indicator of the active

orientation of the field under investigation. The indices revealed patterns of disciplinary

variation that were in concert with those found in Biglan (1973a) and Kolb (1976). Based on

these findings, Kolb (1981) proposed a four-fold typology of academic disciplines. Table 2.1

presents this classification.

Table 2.1. *Kolb's (1981) typology of academic disciplines*

Abstract-Reflective	Abstract-Active	Concrete-Active	Concrete-Reflective
Natural sciences and mathematics	*Science-based professions*	*Social professions*	*Humanities and social sciences*
Mathematics, Chemistry, Physics, Biochemistry, Physiology, Agriculture/Forestry, Zoology, Botany, Bacteriology, Economics,	Mechanical Engineering, Electrical Engineering, Ecology, Civil Engineering, Chemical Engineering,	Education, Social work, Psychology, Law, Educational Psychology, Architecture	English, Political science, Dramatic arts, Journalism, Philosophy, Art, Music, Library science,

As was mentioned above, both Biglan (1973a) and Kolb (1976, 1981) based their classifications on questionnaire data. Becher (1981), however, adopted an interview-based approach to investigate the nature of academic disciplines and disciplinary knowledge. He interviewed 126 academics from six different disciplines, namely, Physics, History, Biology (zoology and botany), Sociology, Mechanical Engineering, and Law. Concentrating on two or three well-regarded departments in each discipline, Becher (1981) conducted interviews with eight members of staff from each department. Becher's interview sample comprised 126 interviews—a minimum of 20 per discipline and a maximum of 24 (in the case of Biology). It is important to note that Becher's (1981) conducted interviews were not structured, that is, the respondents were not presented exactly the same questions, and the answers to the questions were not fixed or closed-ended.

Analyzing the interviewees' responses and comments, Becher (1981) found certain differences and similarities between the selected disciplines. First, he found that some of the distinctions and commonalities between the disciplines rested on their epistemological considerations. For instance, Becher found that most historians share a credo that everything is more complicated than it may appear at first glance. In contrast, engineers tend to take the atomistic view that real-world problems are basically simple and easy to be tackled, if broken down to their component parts. Second, he also found that the role of ideology is rather

limited and negligible in disciplines such as Engineering, Physics and Biology. As Becher contends, this can be due to the heavy reliance of these disciplines on observations, scientific evidence and concrete proof. In History and Sociology, however, individuals' values, beliefs and world views are often reflected in one's approach to investigating a given phenomenon. Third, it was also found that disciplines vary in terms of, what Becher (1981, p. 112) refers to as, "cumulative and critical understanding". As Becher argues, the steady accretion of knowledge in Biology and Physics is maintained by building on previous work; whereas, in History and Law, it is the critical reappraisal and reinterpretation of the less stable body of ideas in these disciplines that lead to the advancement and expansion of knowledge. Finally, Becher's analysis of interviewee's remarks revealed certain differences in the characteristic modes of publication across the six disciplines. For example, it was found that engineers usually limit themselves to journal publication (research papers), whereas historians tend to focus more on publishing books than journal articles. Interestingly, Becher also found that, in Law, case notes are as highly regarded as research papers and scholarly books, and that writing student textbooks can count heavily toward academic promotion in this particular discipline.

Drawing on the above findings, Becher (1981) characterized disciplinary specificity as a continuum between, what he metaphorically refers to as, *urban and rural* research styles. As Becher describes, while urban styles of research tend to focus on discrete and separable problems, rural research often centers on a broad area of enquiry. In other words, urban researchers (e.g., engineers) often go for specific research issues and relatively short-range solutions, whereas rural researchers (e.g., historians) are typically inclined to resolve longstanding issues and tackle broad-based topics. In addition, collaboration is more common in urban that rural research activities, which is the reason why the number of authors involved in conducting a given study in the urban fields (e.g., Physics) is often higher,

compared to the rural disciplines (e.g., Law). Becher's (1981) classification is important as it portrays different research styles and points of view from which members, in a particular disciplinary area, approach various phenomena of interest.

Extending the findings of Becher (1981), Becher (1989) conducted interviews with over 220 academics from 12 disciplines and 18 institutions. Becher (1989) used the patterns of disciplinary variation that emerged from his interviews to bring together the taxonomies of disciplines proposed by previous disciplinary studies, namely, Kuhn (1970), Biglan (1973a) and Kolb (1976b, 1981), and Becher (1981). Akin to these studies, Becher (1989) also adopted bipolar dimensions to provide a framework for distinguishing disciplinary knowledge and place academic disciplines into meaningful categories. Table 2.2 summarizes the characteristics of the disciplinary groupings proposed by Becher (1989) and compares it to the taxonomies introduced in previous studies.

Table 2.2. *Knowledge and disciplinary grouping*

Disciplinary groupings					Nature of knowledge
Kuhn (1970)	Biglan (1973a)	Kolb (1976b)	Becher (1981)	Becher (1989)	
paradigmatic	hard-pure	abstract-reflective	urban	Pure sciences	Atomistic, cumulative, and value-free; consensus over significant questions to address; research activities result in discovery/explanation
paradigmatic	hard-applied	abstract-active	urban	Technologies	Pragmatic and purposive; concerned with mastery of physical environment; criteria for judgement are functional; research activities result in products/techniques
pre-paradigmatic	soft-pure	concrete-reflective	rural	Humanities	Personal, reiterative and value-laden; dispute over criteria for knowledge verification; lack of consensus over significant questions to address; research

					activities result in understanding/interpretation
pre-paradigmatic	soft-applied	concrete-active	rural	Applied social sciences	Utilitarian; functional; concerned with enhancement of professional practice; research activities result in protocols/procedures

As Becher and Trowler (2001) argue, this four-fold typology of disciplinary knowledge can make reasonably clear distinctions between various intellectual fields in academia in terms of: (a) the characteristics of their objects of enquiry, (b) their enquiry procedures, (c) the interaction between the researcher and knowledge, (c) the nature of knowledge accretion and growth, and (d) the results of research. As is evident from Table 2.2, Becher's (1989) classification scheme is a slight yet deliberate modification of the typologies proposed by landmark studies in this area. Altogether, these classifications are important in that they indicate how the characteristics of disciplines and their perceived overlap with other academic areas may affect the scope of their members' scholarly activities.

2.3.2. Linguistic studies of disciplinary specificity

The focus of the above-reviewed studies has been on the epistemologies of academic disciplines and how community-specific theories, methods, and strategies can be used as demarcation criteria for a characterization of disciplinary knowledge. Although the findings from this line of research abound with extensive, thought-provoking comparisons of academic disciplines, they often tend to overlook the importance of language in the construction and transmission of disciplinary knowledge. As Becher and Trowler (2001) argue, the differences and similarities that exist between disciplines can also affect the ways in which knowledge is communicated and presented to its target audience (Becher &

Trowler, 2001). Knowledge in disciplinary communities is negotiated through constructing a form of discourse that not only reflects the collective norms of the community but also adheres to the expectations of its members. Such community-specific meaning is primarily communicated through writing in academia (e.g., Elbow, 1991; Russell, 1991). The use of writing, as the primary means of knowledge production and transmission, in academia is deeply rooted in the institutionalization of the human sciences and the formation of academic disciplines during the Victorian era. According to Russell (1991), before the 1870s, scholarly knowledge was primarily communicated through speaking (e.g., oratory, recitation, debate), and writing was merely an aid to memory and a way of preserving thinking for speech. In parallel with the formation of academic disciplines during the late nineteenth century, however, the predominant role of writing in the production of specialized meaning began to emerge. This was due to the recognition of the need for the transmission of discipline-specific meaning to specialized audiences. As Russell writes, professionals and academics were no longer expected to communicate the outcomes of their scholarly activities to audiences from a wide knowledge base but rather to a group of community members who were united by the shared activities, the goals, and, more importantly, the unique written conventions of their disciplinary community (p. 4).

Commonly described by the term "disciplinary writing", this unique and specific form of language use is characterized by certain linguistic particularities that vary across disciplines. Disciplinary writing can, in other words, represent the linguistic codes that are shared and readily recognizable by community insiders. Such linguistic codes comprise a wide array of rhetorical conventions and lexical and grammatical features (Biber, et al, 1999; Hyland, 2016; Swales. 1990).

Research into disciplinary writing has been guided by two main approaches. One approach has been to study the rhetorical conventions reflected in the discursive and textual

structure of disciplinary texts (Section 2.3.3). A different approach altogether has been to explore the lexical and grammatical features that characterize writing across academic disciplines (Section 2.3.4). Since this dissertation focuses on disciplinary writing, I explore these two approaches in more detail in the following sections.

2.3.3. Structural characteristics of disciplinary writing

This qualitative approach, which was spearheaded by Swales' (1990) seminal work on *genre* analysis, has been mainly focused on the discursive and textual conventions that characterize disciplinary writing and practices. According to Swales (1981, 1984, 1990), genre is a conventionally recognized template through which discourse is constructed with the aim of fulfilling the communicative purposes of social interactions. In academic contexts, genres enable writers to communicate their message in a way that is retrievable by the reader. The communicative purposes of such 'writer-reader' interactions are realized through rhetorical choices that instantiate the expectations of the target readership in a particular academic area. These rhetorical choices mainly involve decisions regarding the overall organization of the discourse, as well as the linguistic resources employed to reflect their communicative purposes (Bhatia, 1993).

Adopting this approach, several genre-based studies to date have investigated the rhetorical and organizational structure of different sections of research articles in a range of academic disciplines. For example, Samraj (2002) conducted a genre analysis of research article Introductions from two related disciplines: Conservation Biology and Wildlife Behavior. She examined the textual characteristics of the Introduction section of 24 research articles published in key journals in the two fields. The results of her analysis revealed that the Conservation Biology Introductions fulfill a greater promotional function than the Wildlife Behavior Introductions. She also found that Introductions written in Wildlife

31

Behavior tend to focus on the epistemic and theoretical aspect of the research being reported, as against the Conservation Biology Introductions which were found to impart a real-world and laboratory sense. Samraj's (2002) findings are important because they indicate that, by comparing the same text types across different disciplines, we can distinguish textual characteristics that are reflective of disciplinary norms. Extending Samraj's (2002) findings, Samraj (2005) investigated the rhetorical characteristics of disciplinary writing in two different text types: research article Abstracts and Introductions. Focusing on the same two disciplines (i.e., Conservation Biology and Wildlife Behavior), she carried out a genre analysis of the Abstract and Introduction sections of 48 research articles. The results of the analysis showed that Abstracts and Introductions are interrelated in different ways in the two disciplines. Specifically, she found that Abstracts and Introductions in Conservation Biology bear a greater similarity in terms of their communicative functions and textual organizations, compared to Abstracts and Introductions written in Wildlife Behavior. Samraj's (2005) results illustrate that disciplinary writing may manifest itself differently across different sections of research articles.

In another study, Bruce (2008) conducted a genre-investigation of Methods sections reporting research in the physical (Biology, Organic chemistry, Medicine, Chemical Engineering) and social sciences (Applied Linguistics, Education, Sociology, Psychology). Analyzing a corpus of 60 research Methods (30 from each of the two disciplinary domains), Bruce found Methods sections in the physical sciences to be compact and dense due to the extensive use of noun phrases by writers in these fields. He also found that writers in the physical sciences tend to employ prepositional phrases to describe research procedures and processes (e.g., the nuclei were released according to the method of Sgorbati et al). In contrast, social sciences writing in this section was found to be explicitly interpretative and discursively elaborate, which was reflected by the extensive use of *that* clauses. Taken

together, Bruce's (2008) genre-based investigation of disciplinary writing in Methods sections of research articles reveals fundamental differences in the ways in which writers from across different disciplines present their research methods.

Focusing on another key section of research articles (i.e., the Discussion section), Holms (1997) used an existing model of discourse organization based on natural sciences (proposed by Hopkins & Dudley-Evans, 1988, p. 118) to explore variation in discussions of results in three social sciences disciplines: Political sciences, Sociology and History. He examined the rhetorical structure of 10 discussions (or conclusions) from each of the three disciplines. Holms found that, of the three disciplines, the History Discussions bore the least resemblance to those of the natural sciences, and Political and Sociology texts closely resembled each other but were also sufficiently distinct from natural sciences texts in terms of brevity and complexity. In a similar study, Basturkmen (2012) examined the rhetorical structure of Discussion sections in Dentistry with reference to a schematic framework of discussion sections in Applied Linguistics (outlined by Basturkmen, 2009). The rationale for choosing Dentistry, as Basturkmen (2012) explains, was that published information about research writing by dentists is very rare and that more research is needed to shed light on common writing practices in this particular discipline. Examining Discussion sections of ten research articles in the field, she observed that providing a detailed explanation of findings was less common in Dentistry, in comparison to the Applied Linguistics Discussions. Basturkmen also found that the Discussion sections in Dentistry generally opened with some background information, whereas social sciences Discussions typically opened with statements of results. Another interesting pattern of results also showed that writers in Dentistry often tend to combine two complex forms of argumentation (i.e., comparison and evaluation of results) to construct elaborate cases for the reliability of their findings, a rhetorical strategy which was found to be less commonly used in social sciences Discussions.

Both Holms (1997) and Basturkmen (2012) provide important insights into how disciplinary specificity can influence the rhetorical strategies scholars employ to discuss the results of their research.

In another genre-based study focused on disciplinary writing, Stoller and Robinson (2013) investigated all the four key sections (Introduction, Methods, Results and Discussion/Conclusion) of research articles in a single discipline: Chemistry (as an under-researched discipline). Stoller and Robinson (2013) examined approximately 60 full-length research papers from six accredited journals in the field. Their analysis revealed several interesting findings concerning the organizational structure of Chemistry research papers. First, it was found that Chemistry journal articles often do not include Literature review sections and that the Methods section (commonly labeled 'Experimental or Materials and Methods') is placed immediately after the Introduction. Their analysis showed that writers in Chemistry tend to incorporate their references to the literature into Introduction and Discussion sections. Second, Stoller and Robinson also observed that Results and Discussion (R&D) sections in Chemistry research papers tend to fall on a continuum between completely separated and fully merged R&D sections. More specifically, four patterns of R&D sections emerged: (i) fully separated R&D, (ii) blocked R&D – the sections were not either completely separated or fully merged, such that a single block of results was commonly followed by a block of discussion, (iii) iterative R&D -- in this pattern, authors were found to alternate between presenting and discussing results, and (iv) fully integrated R&D -- results were presented and discussed seamlessly.

Stoller and Robinson (2013) also compared their results to those of Kanoksilapatham (2005, 2007), who also conducted a genre analysis of 60 full-length research articles published in Biochemistry--a sub-discipline of both Chemistry and Biology. Stoller and Robinson's comparisons revealed certain similarities and differences between these two

neighboring disciplines. First, Introductions in both fields included a section that served as a transition to the study and its purpose. Further, neither Chemistry nor Biochemistry Introductions concluded with outlines of the remainder of the article. The comparison of the same section in the two disciplines also revealed that the Biochemistry Introductions included a preview of research findings, a rhetorical strategy which was found to be rarely present in the Chemistry Introductions. Second, both Chemistry and Biochemistry Methods sections included a detailed description of instrumentation (e.g., spectrophotometers, mass spectrometers). In both disciplines, authors were found to include only essential information in the description of their methods and often referred the reader to the related literature. Third, both Chemistry and Biochemistry Results sections were found to open with a reminder of how the results were obtained. This reminder was rather extensive in Biochemistry, compared to Chemistry. Interestingly, Results sections in Biochemistry also included a justification of procedures and methodology. This rhetorical strategy was found to be completely absent in Chemistry Results. And finally, Discussion sections in the two disciplines were relatively similar, although Biochemistry writing in this section was found to include statements about the limitations of the reported study and suggestions for future research – two communicative strategies which were absent in Chemistry Discussions. In sum, Stoller and Robinson's (2013) comparison of findings with those of Kanoksilapatham (2005, 2007) is important in that it illustrates the diverse nature of disciplinary writing in key sections of research papers.

As is evident from this brief, and no doubt selective, review of research, genre-based investigations of disciplinarity can provide important insights into how research writing, as the primary means of knowledge dissemination in academia (e.g., Bazerman, 1994; Hyland, 2016; Flowerdew, 2013b), can be affected by disciplinary conventions. Genre-based studies, however, have been criticized for their unsystematic, and mainly qualitative, approach to

studying academic texts (Cortes, 2015; Crawford & Csomay, 2015; Hyland, 2013; Paltridge, 1994). Due to the qualitative nature of this approach, researchers often resort to analysing a small number of academic texts which, in turn, limits the generalizability of their findings (Biber, Connor, and Upton, 2007, p. 36). In addition, genre-based studies have mainly focused on aspects of discourse beyond the sentence level, generally disregarding lower-level linguistic characteristics such as lexical items and grammatical features of texts. As Dudley-Evans (1994) argues, decisions regarding the discourse structures of research articles, or other academic genres, are mainly made on the basis of their associated linguistic features. As Haswell (1991) points out, different genres have different preferences for the phraseology used to fulfill their communicative purposes. That is, the reason why specific genres, such as 'lab reports', can be clearly distinguished from other narrative genres, such as 'fiction', is due to the linguistic features that are distinctive of their style of communication. These linguistic resources can take the form of both single and multi-word items. Lim (2006) investigated the lexical and syntactic devices that are commonly used to fulfill the communicative purposes of Method sections of Management research articles. Lim found that authors use certain linguistic clues, both in the form of single and multi-word units, to indicate internal boundaries in their texts. His findings also revealed that the majority of these single word items were co-occurring with certain words, helping to convey the overall communicative aims of the Methods section. Lim's (2006) findings show that the use of particular linguistic (lexical and grammatical) features can be closely tied to the communicative functions/purposes of a text. In the following section, I explore the studies that have focused on investigating such lexical and syntactic features in disciplinary writing.

2.3.4. Lexical and grammatical characteristics of disciplinary writing

As was mentioned in Section (2.3.2), a different approach to studying disciplinary writing has been to investigate the frequently used lexical and grammatical features that distinguish academic discipline and their conventional writing practices. This approach, which was made possible by advances in corpus methodologies in the late 1970s, focuses on linguistic variation in a large and principled collection of texts across various disciplines. Applying this approach, a multitude of corpus-based studies have explored disciplinary variation by looking at single word vocabulary (e.g., Hyland & Tse, 2005; Durrant, 2014; Jiang & Hyland, 2017), multi-word expressions (e.g., Cortes, 2004; Durrant, 2017; Hyland, 2008; Omidian, Shahriari, Siyanova-Chanturia, 2018), and overall lexico-grammatical patterns of language use (e.g., Afros & Schryers, 2009; Biber, 1988; Biber & Finegan, 2001; Conrad & Biber, 2001; Gray, 2016) across disciplines in academic texts. This large body of research can also be categorized into two major research areas: (1) studies that have investigated these linguistic features within a single academic discipline and (2) those that focus examined these features across multiple disciplines. In what follows below, I present a survey of previous studies in these areas, documenting the linguistic characteristics of academic (research) writing.

2.3.4.1. Disciplinary specificity in single-word academic vocabulary

The main aim of research in this area has been to highlight the mediating role of disciplinary specificity in the use of academic vocabulary. The term academic vocabulary refers to lexical items (single-or multi-word items) commonly used to construct knowledge in spoken and written academic registers (e.g., lectures, seminars, research articles, book reviews). This vocabulary is typically classified into two sub-categories: general and technical academic vocabulary (Hierber & Lubliner, 2008; Nation, 2001). General academic vocabulary refers to

those lexical items that are not tied to a specific discipline and are commonly used in a wide variety of scientific disciplines (Coxhead, 2000). This vocabulary, which includes words and sequences such as *insight*, *exhibit*, *as a result of*, falls somewhere between non-academic and technical vocabulary (Beck, McKeown, & Kucan, 2002, 2008; also see Coxhead, 2019).

Technical (or discipline-specific) vocabulary is a set of subject-related words or word sequences (e.g., *estrogen, ultra high-energy cosmic rays*) that are used to create specialized knowledge in a given domain of enquiry (Chung & Nation, 2004). Such domain-specific lexes are used by members of a particular academic community to construct a specialized form of discourse which is exclusive to their target audience and may not be readily understood by members of other disciplinary domains. However, as Liu and Lei (2019) argue, there are two views on the scope of technical vocabulary. On the one hand, there is a broad view that places technical vocabulary on a continuum that includes words of various types, ranging from those that are almost always exclusive to a particular subject area (e.g., *oxidative* in Chemistry or *periodontal* in Dentistry) to those that might boast a high frequency of use in general English yet convey domain-specific meaning (e.g., *tension, stress* in Mechanical Engineering or Applied Physics) (e.g., see Nation, 2013; Nation & Coxhead, 2012). On the other hand, there is a narrow view that assumes a distinction between technical and specialized vocabulary. According to this view, the former vocabulary is confined to scientific terminologies commonly found in hard knowledge fields (e.g., Biology, Chemistry, Mechanical Engineering), while words belonging to the latter category are those that are often used in Arts and Humanities and Social sciences (i.e., soft knowledge fields) (e.g., see Brieger & Pohl, 2002). As Liu and Lei (2019) reasonably argue, making such a distinction is not particularly meaningful or feasible for the definition of technical vocabulary, as it only adds to the confusion of terminologies used in the discussion of this vocabulary. Furthermore, with the increase in interdisciplinary work in academia, many scientific research/statistical

methods (e.g., mixed effects modeling), and their associated technical terminologies, have become commonplace in soft knowledge fields; and thus, restricting the scope of this vocabulary to a particular domain of enquiry does not appear to be an effective approach (ibid). For these reasons, the present dissertation takes a broad view on the scope of technical vocabulary and does not confine the definition of these words to a particular disciplinary domain.

The rationale for investigating disciplinary specificity in the use of academic vocabulary has rested on the argument against the idea of 'generality' in academic discourse (Blue, 1988; Jordan, 1997). In essence, the idea is related to the assumption that particular areas of academic discourse are relatively homogenous in nature, such that identifying and listing the linguistic features (e.g., academic words) representing this homogeneity would be facilitative to teaching English for Academic Purposes (Hutchinson & Waters, 1987, p. 165-166). The argument against such a coarse-grained approach to understanding academic discourse has two prongs: (a) some words are likely to be more important for academic (research) writing in certain disciplines, compared to others (e.g., Durrant, 2016; Hyland & Tse, 2007; Martinez, Beck, and Panza, 2009) and (b) it is highly unlikely that these words are evenly dispersed across disciplines, and thus they may not be equally useful for writing in all branches of academic study (e.g., Durrant, 2014). In line with this argument, Hyland and Tse (2007) investigate disciplinary variation in the use of academic single words in a range of academic disciplines. For this purpose, Hyland and Tse used Coxhead's (2000) Academic Word List (AWL) to examine disciplinary variation in the frequency of occurrence and meaning of the items from the list in a multidisciplinary corpus of academic texts (3.3 million words). Coxhead's (2000) influential AWL consists of 570 sub-technical word families (i.e., headwords and their inflectionally- and derivationally-related forms) frequently used in academic writing across disciplines. The AWL was compiled from a 3.5 million-word corpus

of academic texts written in a wide range of disciplines. Examining disciplinary variation in the use of AWL items, Hyland and Tse (2007) found that the list offered the lowest coverage for science writing (9.1%) in their corpus. This finding is also supported by those of Coxhead (2000), who also found that her list covered the lowest number of words in academic texts written in the sciences. This variation may suggest that, as Hyland and Tse (2007) argue, science writing relies heavily on more specialized and technical vocabulary, which might be particular to this specific disciplinary area. Further, the authors also found that individual items on the AWL often have different meanings and occur in different collocational environments, depending on the disciplines in which they are used. For example, it was found that the word *process* (one of the most frequent AWL items) is highly likely to occur as a noun in sciences and engineering texts. This was taken as an indication of greater reliance on *nominalization* (more on this later) in hard sciences writing, as opposed to writing in the soft knowledge fields. In another example, Hyland and Tse also observed that the word *analysis* often tends to occur in discipline-specific compound nouns, such as *genre analysis* (Applied Linguistics) and *neutron activation analysis* (Physics). It can be argued that producing discipline-specific meaning using such technical expressions demands extensive, specialized knowledge of all the words forming these phrases, and not just the word *analysis.* Based on these findings, Hyland and Tse (2007) argue that it is difficult to identify uniformity in academic vocabulary use and that further research is needed to shed light on various parameters of variation in the use of academic words and word sequences. Hyland and Tse's study provide a clear indication that disciplinary practices not only operate at the macro levels, such as discipline-specific epistemologies and community-based methods and strategies, but also involve variation at the level of language and vocabulary use.

In a similar vein, Martinez, et al. (2009) explored disciplinary specificity in academic vocabulary use in a corpus of research articles (826,416 words) in the agricultural sciences.

Integrating corpus-based and genre-based approaches (see Sections 2.3.3. and 2.3.4), Martinez and colleagues investigated instances of such specificity across the four main sections of the research article (IMRD). Using the AWL as a match list, the authors identified patterns of variation across the four sections. The results of their analysis showed considerable variation in the use of academic word types (individual words) across sections, with Results and Discussion sections containing the lowest and the highest number of word types, respectively. As the authors argue, the highest number of word types used in the Discussion section can be attributed to the argumentative and discursive nature of this section of research articles. The authors also observed considerable variation in the number of word families (i.e., a collection of word types bound together in terms of form and meaning, see Bauer & Nation, 1993) across the four sections, with the Introduction containing the highest number of word families. The authors compared the 20 most frequent words in the four sections of agricultural sciences articles with the 20 most frequent words in the medical corpus of Chen and Ge (2007), who also investigated disciplinary specificity in academic vocabulary use across the IMRD sections of medical research articles. The comparison of the same sections in the two sciences (i.e., agriculture and medicine) revealed a higher degree of overlap (in the use of frequent academic words) between Results and Discussion than Introduction and Methods sections. Further, Martinez, et. al.'s (2009) observations also revealed that the high-frequency words (e.g., *show*) that are often excluded from academic words lists conveyed academic meaning and played an important role in the construction of arguments across sections of research articles. Similar to Hyland and Tse (2007), they also observed that these non-academic words can carry different meaning senses, and take on specific contextualized meaning, depending on the collocational environments in which they commonly occur (e.g., *control* in *control group* vs *biological control agent*s). Based on these

observations, Martinez et al. (2009) suggest that words should not be excluded from an academic word list on the grounds that they are prevalent in general (non-academic) English.

In light of the findings from Hyland and Tse (2007) and Martinez et al. (2009), Durrant (2016) set out to investigate the academic words commonly employed in university students' writing, using the items from the Academic Vocabulary List (Gardner & Davies 2014) as a reference list. The Academic Vocabulary List (AVL) consists of 3,014 lemmas (i.e., headwords and their inflectionally-related forms), frequently used in the academic sub-corpus of Corpus of Contemporary American English (COCA, Davies, 2009). This sub-corpus of COCA contains 120 million words of texts from across nine disciplines published in various academic journals, academically oriented magazines, and the finance sections of newspapers. To evaluate the extent to which university student writing draws on the AVL in their writing, Durrant (2016) compared the list with lemma frequency lists derived from the British Academic written English (BAWE) corpus (Nesi, Gardner, Thompson, & Wickens, 2008-2019). The BAWE corpus comprises a collection of 'successful' course assignments (c. 3000 texts) produced by Bachelors- and Masters-level students at different universities in the UK. Assignments in BAWE are deemed successful as they all had received at least an 'upper-second class' or 'merit' grade (comparable to a 'B' grade in the US system) when assessed by subject area tutors (Nesi & Gardner, 2012). All texts in the BAWE corpus are marked for assignment type, academic level, and discipline, among other things. In addition to evaluating AVL's coverage of words in BAWE, Durrant (2016) also examined variation in the use of the list across student level, text type, and disciplines. The results of his analysis showed that the AVL offers a good overall coverage (c. 34%) of lexical words in BAWE texts. This coverage was also found to increase as a function of students level of study (e.g., 34% coverage for Masters level assignments vs 29% for first-year undergraduate texts), suggesting that the AVL may be a valuable resource for student writing in tertiary education.

However, Durrant (2016) observed that AVL's coverage significantly varied across text types and disciplines. While variations across text types were relatively minor, disciplinary differences were found to be rather extensive--with AVL's mean coverage ranging from 21% to 39% across disciplines. Durrant's (2016) analyses also revealed that students from certain disciplines (e.g., Classics) in BAWE make much more intensive use of a small number of AVL items, compared to their peers in other disciplines (e.g., Architecture). Based on these observations, Durrant (2016) concluded that disciplinary-specificity is a more important driver of variation in university student writing, in comparison to other situational variables such as text type (also see Durrant, 2017, Omidian, Siyanova-Chanturia, & Durrant, 2020).

The evidence from this line of research seems to endorse the idea of specificity in academic vocabulary use. However, since these studies mainly focus on the evaluation of vocabulary lists (i.e., the AWL and the AVL), it can be argued that their conclusions about specificity in academic vocabulary use are largely restricted to the items included in these lists. The approach adopted in these studies is to use corpus evidence to analyze different patterns of use for a set of pre-defined items. In other words, corpus analysis in such studies is viewed as an evidential tool to validate theoretical presumptions about a collection of linguistic items. However, more generalizable conclusions can be reached by adopting an inductive approach in which patterns of disciplinary specificity in vocabulary use emerge from the analysis of a corpus with minimal a priori assumptions guiding their identification (more on this later). Adopting a similar approach, Durrant (2014) examined the degree to which vocabulary use remains constant across different groups of university students, without focusing on a particular pre-defined list of items. To this end, Durrant (2014) first created separate listings of high-frequency vocabulary for BAWE texts across different disciplines and at different levels of study. Words were considered highly frequent if they occurred (i) more than 100 times per million words and (ii) if they were used in more than 10% of all

texts in the corpus. It is also important to note that, unlike previous studies, Durrant (2014) focused on word forms rather than more abstract categories such as word families and lemmas. As he argues, since different forms of a word can have markedly different frequencies across academic disciplines, conflating word forms into lemmas or word families runs the risk of disguising important patterns of variation.

Once the lists of high-frequency word forms were retrieved from the corpus, Durrant (2014) then used Gries's (2008) deviation of proportions (DP) statistic to investigate the extent to which these frequent word forms were evenly dispersed across disciplines and levels of study in university student writing. The results of this analysis revealed that the distribution of high-frequency words was considerably uneven across disciplines. Durrant (2014) also observed that discipline-specific words form a substantial portion of vocabulary use in certain disciplines (e.g., Agricultural Science) and that only half of the high-frequency words that are important for particular disciplines are generic to the writing of other subject areas. Following this analysis, Durrant (2014) also examined the degree of overlap between the lists of high-frequency words. Overlaps were quantified as the percentage of high-frequency words that are shared by two disciplines. A hierarchical cluster analysis was then conducted based on these overlap percentages to determine how discipline might group together in terms of their academic vocabulary use. It was found that students at different levels of study in a discipline tend to be relatively homogeneous with regard to their vocabulary use. Durrant (2014) also found that vocabulary use across disciplines can be classified into four main clusters: Humanities/Social Sciences, Business/Economics, Social/Health Sciences, and Science/Technology. Based on these observations, Durrant (2014) concluded that exploring the variation in language use across disciplines can provide a solid basis for decisions concerning the choice of materials and design of academic programs.

Taken together, the above-reviewed studies show that an awareness of disciplinary differences in the use of academic vocabulary can be a foundational tool for developing teaching materials for EAP applications (also see Biber & Conrad, 2009, p. 269).

2.3.4.2. Disciplinary specificity in multi-word academic vocabulary

It has long been acknowledged that certain words tend to co-occur in specific configurations (Biber, et al., 1999; Schmitt, 2004; Wray, 2002). These configurations are often referred to as *multi-word expression*s (MWEs), a terminology that encompasses a wide range of linguistic units above the single word level (as opposed to *formulaic language* which may also include single word items, Van Lancker-Sidtis & Rallon, 2004; Wray, 2002; also see Siyanova-Chanturia & Omidian, 2020 for a discussion). MWEs can take the form of collocations (*strong coffee*), binomials (*bride and groom*), idioms (*spill the beans*), lexical bundles (*on the one hand*), and others. According to some estimates, these sequences comprise somewhere between 20 to 50 per cent of natural language production (Biber et al., 1999; Erman & Warren, 2000; Foster, 2001; Howarth, 1998; Sorhus, 1977).

In the past three decades, considerable attention has been paid to MWEs as teachable units for EAP/ESP applications. More specifically, due to their pervasive role in language use, MWEs are believed to be of central importance in constructing and shaping meanings in academic prose. Evidence suggests that these multi-word items can also be closely linked to particular communicative functions and play a key role in the construction of discipline-specific meaning in a register (e.g., Cortes, 2013; Durrant, 2017; Hyland, 2008a, 2008b; Le & Harrington, 2015; Omidian, et al, 2018). As a result, EAP/ESP course instructors are advised to not only raise students' awareness of the organizational structures of texts and their single-word linguistic realizations, but also compliment this instruction with teaching the multi-word items through which discipline-specific meanings are commonly created and

communicated in key parts of academic prose (e.g., the Introduction section of research articles; see Cortes, 2013).

Appropriate use of MWEs can lead to both authentic and predicted communication (Li & Schmitt, 2009; Schmitt & Carter, 2004). Expert writers in most discourse communities tend to make frequent use of certain pre-fabricated patterns (e.g., *as can be seen*) in their production and expect to see the same expressions in the texts they themselves encounter (Nesselhauf, 2004). Such sequences of words are often seen as linguistic shortcuts that assist the communication of meaning by making language more predictable and familiar to the reader or hearer (e.g., Wray & Perkins, 2000; Nattinger & Decarrico, 1992). In fact, language users, native or nonnative, who fail to make use of such characteristic features may come across as being insufficiently experienced in the target register (Haswell, 1991; Hyland, 2008). In this case, they would need to 'proximate' themselves to the linguistic conventions of the target register and learn how to use these conventions to negotiate their ideas in a way that is both accepted and expected by other community members (Hyland, 2015). This is particularly true for academic research writing, through which researchers from various disciplinary communities seek to disseminate their scientific findings through publishing their work in English. And thus, there are advantages to studying MWEs in academic (research) writing to better help students, and early career researchers, acquire the specialized discourse competencies required to participate in research and publishing activities in their fields.

Traditionally, the study of MWEs was mainly based on intuition, with researchers relying on rather impressionistic evidence for their identification (e.g., Cowie 1994; Howarth 1998; Foster, 2001). This approach is often referred to as a 'phraseological' (or qualitative) approach to defining and identifying MWEs. Very loosely, the term phraseology can be defined as "the study of the structure, meaning and use of word combinations" (Cowie, 1994,

p. 3168). The focus of this traditional approach was mainly on distinguishing compositional phrases from non-compositional ones, as well as making intuitive judgments regarding their semantic properties (e.g., literal vs figurative).

Following the developments introduced by corpus-based methodologies, however, a more systematic approach emerged in which frequency and distributional properties of sequences are deemed to take precedence over intuition and perceptual judgments (e.g., Sinclair, 1991). By utilizing large text corpora and various statistical metrics, such as frequency and measures of association strengths (mutual information, t-score), proponents of this approach have developed several methodologies to study the use of MWEs in natural language. Two general approaches that have emerged from these efforts are: 'corpus-based' and 'corpus-driven' (Tognini-Bonelli, 2001). In the corpus-based approach, corpus evidence is used to analyse different patterns of use for a set of pre-defined sequences. This approach relies heavily on theoretical presumptions regarding the semantic and syntactic properties of MWEs and uses corpus analysis as an evidential tool to validate these pre-established conceptions (Erman & Warren, 2000; Moon, 1998; also see Section 2.3.4.1). The corpus-driven approach, however, is more inductive (bottom-up) in nature, in the sense that sequences emerge from the corpus analysis with little theoretical assumptions guiding their identification (Biber, 2009). In this approach, it is the emerged linguistic patterns that form the foundation of a linguistic theory, and not vice versa.

More recently, by adopting a bottom-up (corpus-driven) approach, Cortes (2013) examined the use of lexical bundles in the moves and steps of research article introductions. She found a strong connection between the communicative functions of certain bundles and the rhetorical aims of the moves and steps in which they were found. Cortes (2013) also observed that certain bundles were used by authors to signal the onset of a move and/or step. Similarly, Le and Harrington (2015) investigated the use of MWEs in the Discussion section

of quantitative research articles from the field of applied linguistics. They too found that the use of specific word combinations strongly manifested both the genre and the communicative aims of the moves in which they were used. Mizumoto, Hamatani, and Imao (2017) explored the pedagogical implications of the correspondence between lexical bundles and moves by developing a web-based tool that allows the user to search for the most frequently used bundles in different moves of applied linguistics research articles. The authors piloted the tool with eight LX speakers of English for two to five months. Their results showed positive user feedback on the pedagogical implications of the tool and the bundle-move approach for academic writing. In a similar vein, Omidian et al (2018) explored disciplinary writing, as reflected in in the use of MWEs in different rhetorical moves of research article abstracts. Adopting a corpus-driven approach, the authors identified the most frequently occurring n-grams (i.e., lexical bundles) of different length in a corpus of around 6000 abstracts from six disciplines. The identified bundles were then classified according to their communicative functions in different moves of the abstract. Omidian, et al (2018) found that members of different academic domains have different priorities for representing their research in academic abstracts. More specifically, it was observed that writers of abstracts in the soft sciences tend to ensure that the objective of their research and its contribution to the field is clearly expressed. On the other hand, writers of abstract in the hard sciences see it as crucial to promote their research through the lens of its methodology and apparatus.

These studies are important because they provide evidence for the strong connection that exists between MWEs and communitive purposes of different parts of research articles, and argue against traditional approaches towards discourse analysis in which the identification of discourse functions and their linguistic realizations were solely based on qualitative evidence drawn from subjective observations (see Biber, Connor, & Upton, 2007, p. 36). What is largely missing in the described studies, however, is a systematic analysis of

the 'form-function' link across sections of research articles in different disciplinary fields

(see Chapter 5). Specifically, more empirical evidence is needed to develop a better

understanding of disciplinarity through the lens provided by the interaction between form and

function in academic texts (see Moreno & Swales, 2018).

Knowledge of academic word sequences plays an important role in the construction

of meaning in written academic genres. Many studies to date have explored the use of

phraseological patterns in a wide range of academic texts. Some prominent studies on this

topic have included: textbooks (Coxhead, Stevens, Tinkle, 2010; Konstantakis, 2007), student

essays (Durrant, 2014, 2016; Adel & Erman, 2012), doctoral dissertation and master's theses

(Hyland, 2008a, 2008b), book reviews (Groom, 2009; Romer, 2010). As was mentioned

earlier, one of the primary means of disseminating knowledge in academic contexts is

research articles. Research articles are regarded as a highly conventionalized genre in which

effective transmission of knowledge is heavily dependent on writers' familiarity with the

linguistic expectations of their target readership (Bhatia, 1993; Hyland, 2001; Swales, 1990).

Such linguistic expectations are mostly related to the use of academic words and sequences

that are conventionally accepted and expected by readers in a given disciplinary community.

Using domain-specific corpora, a multitude of studies have investigated the use of such

linguistic resources in research articles from across different disciplines (e.g., Cortes, 2004;

Hyland & Tse, 2007; Lee & Harrington, 2016; Martinez, Beck, & Panza, 2009; just to name a

few). Findings from these studies indicate that the use of certain academic words/sequences

and their communicative functions can reflect the epistemological orientations of academic

fields in the construction of knowledge in research articles. Hyland (2011) demonstrated that,

in order to construct a plausible argument in high-stakes genres such as the research article in

which effective writing involves the anticipation of alternative interpretations and readers'

possible objections, authors often tend to make use of linguistic resources that reflect the

disciplinary values and expectations of their target readership. These findings, together with the increasing attention to publication in the form of journal articles in academia (Flowerdew, 2013), highlight the importance of investigating the MWEs that are commonly used by authors in different academic fields to construct knowledge in research articles.

2.3.4.3. Disciplinary specificity in lexico-grammatical patterns

It has long been argued that meaning in language is created as a result of the interaction between lexis and grammar (Biber et al., 1999; Sinclair, 1991; Halliday, 1989; Hunston & Francis, 2000). Research has shown that this interaction is neither arbitrary nor coincidental; rather, it is systematically linked to specific purposes of any given communicative situation. For instance, Biber and Gray (2010) showed that, due to the informational purpose and compact discourse style of academic prose, academic writing tends to rely on the use of nouns, nominal phrases and embedded phrasal structures (e.g., prepositional phrases as post-modifiers, *the observed post-test crack patterns **for** the concrete blockwork and clay brickwork walls*; more on these in Chapters 5 and 6). A prototypical example of this discourse style is the extensive use of nominalization, a process through which meanings are decoupled from their verbal and adjectival realizations and reconstructed using nominal forms (e.g., *optimize, optimization, nominalize, nominalization*, see Halliday, 2004; Biber & Gray, 2016). Biber and Gray (2010) findings also demonstrated that the nominal and phrasal discourse style of academic writing stands in sharp contrast to the clausal discourse style of real-time conversation, which tends to heavily rely on clausal structures (especially *that*-and *WH*-complement clauses and *if-* and *because*-clauses functioning as Adverbial clauses) and verbs (also see Biber, 1988). The following excerpts illustrate how the interaction between lexis and grammar differs between academic writing and real-time conversation. For

reference, head nouns are *italicized*, prepositions are in small caps, verbs are **bolded**, phrasal structures are <u>underlined</u>, and clausal structures are presented in [brackets].

> Coverage *change* <u>of a certain flowering functional group</u> **was** the *difference* <u>between coverage at receiver and donor sites</u>. Temperature *sensitivity* <u>of the phenological events (d/C)</u> **was calculated** as <u>the *difference* in the timing</u> <u>of phenological events</u> [**divided** by soil temperature *change* (C) <u>between the receiver site and donor site</u>]. (extracted from the corpus used in this research, see Chapter 3)

> Yes, [if I'm not **using** the *grill*] I **take** it out. It would **take** a long *time* [to **get** our *grill* out [because I don't **think** (*that deletion*) [it would **move** any more]]]. (extracted from the spoken subsample of the BNC, source information: KC4, S_conv)

The above examples show that academic research writing has distinct linguistic characteristics that set it apart from other language registers such as conversation. As was noted above, these distinguishing characteristics are inextricably linked to the ways in which the lexical and grammatical features of language work together to create meaning and serve various communicative functions in academic (research) writing.

Research in this area has focused on investigating particular lexico-grammatical patterns that can characterize disciplinary writing as a sub-register of academic writing. The rationale behind this research agenda stems from the premise that certain lexico-grammatical patterns (also referred to as *pattern grammar*, see Hunston & Francis, 1999) serve particular communicative purposes, which are often characterized by community-specific practices. In line with this aim, Groom (2005) investigated the use of 'introductory *it*' across two academic disciplines: Literary Criticism and History. Introductory *it* constructions are lexico-grammatical patterns in which the dummy pronoun *it* is followed by a copula/linking verb (*be*, *become*, *seem*, *appear*), an adjectival predicate, and complement clauses (e.g., finite or

non-finite *that*-clause, *to*-clause, *WH*-clause, *it is worth mentioning that*, see Biber, et al., 1999, p. 660 for a discussion). Groom (2005) classified all occurrences of introductory *it* constructions into six categories based on their communicative functions in context: adequacy (e.g., *it is insufficient to argue*), desirability (e.g., *it would be interesting to know*), difficulty (e.g., *it is difficult to imagine*), expectation (e.g., *it is not surprising to find*), importance (e.g., *it is significant that*), validity (e.g*., it is clear that*). Groom (2005) found fundamental disciplinary differences in how writers from the two disciplines employed introductory *it* patterns to persuade readers to accept their interpretations. More specifically, it was found that introductory *it* forms denoting validity and evidential obviousness (e.g., *it is ADJ that*, *it is clear that…*) were more common in Literary than History articles. Groom's (2005) results also showed that *it would be ADJ to-inf* (e.g., *it would also be erroneous to characterize*), which are often employed to deal with anticipated reader objections, were used differently across the two disciplines. In particular, while writers from Literary Critics used these lexico-grammatical patterns to persuade readers to accept extreme interpretations or dissenting views (e.g., *while it would be naïve to identify*), Historians employed these patterns to persuade readers to reject plausible alternative interpretations (e.g., *it would be wrong to see them as…*). Groom's findings demonstrate that the communicative functions of certain lexico-grammatical patterns vary in systematic ways across disciplines.

In a similar vein, Hyland and Tse (2005) examined disciplinary variation in the use of *that*-complement clauses functioning as an 'evaluation' device in research article abstracts from six disciplines. 'Evaluative *that'* complement clauses are often seen as stance-taking features that allow the writer to present an 'explicit' statement of evaluation about a proposition (see Hyland, 2005). The authors found that evaluative *that* clauses signaling stance were virtually always expressed as an epistemic (and not attitudinal) evaluation (e.g., *we demonstrated that, it is highly likely that*). Classifying epistemic evaluative *that*-clauses

based on their specific communicative functions, Hyland and Tse (2005) found that writers in the hard fields often tend to use these lexico-grammatical patterns to express a convincing degree of certainty and assurance about their propositions. This discourse strategy was particularly evident in the extensive use of evaluative *that* clauses controlled by 'factive' verbs (e.g., *demonstrate, prove*) that impart a higher degree of certainty to the following clause than do 'non-factive' verbs, such as *argue, suggest*. In contrast, writers in soft knowledge fields were found to use evaluative *that*-clauses to express their evaluations of a given proposition in a less direct and assured manner. Hyland and Tse (2005) observed that the epistemic evaluative *that*-clauses used in Humanities and Social sciences were dominated by tentative (non-factive) verbs that were employed to mark propositions as being a suggestion, argument or assumption (e.g., *we believe, however, that; these patterns of results suggest that*).

Hyland and Tse's (2005) were further corroborated by Charles (2006), who also found that stance taking is systematically linked to both lexical and grammatical patterns of use across disciplines. Comparing the use of verb-controlled clauses to report information in theses written in Politics and Materials science, Charles (2006) observed that the use of *that*-clauses controlled by factive verbs was more common in Materials science than in Politics. Her findings also revealed that the majority of *that*-clauses employed in Materials science were extraposed clauses realized through a dummy *it* and passive verb (e.g., *it was demonstrated that*). Extraposed clause constructions report a stance which is not overtly attributed to the author, concealing the writer's attitude about the proposition encoded in the complement *that*-clause (see Biber, et al. 1999). Moreover, as Charles (2006) argues, the use of passive tense also allows the writer to obscure their role in the scientific process and instead prioritize the role of research. This emphasis on minimizing the role of the researcher and constructing impersonality and objectivity may reflect something of a scientific ideology,

in which research findings are often anchored in empirical observations, rather than theoretical interpretations (more on this in Chapter 4, Section 4.4). In a similar study, Charles (2007) examined disciplinary specificity in the use of *'noun-that'* patterns (e.g., *the assumption that there is a…*) in the same corpus used in Charles (2006). In *'noun-that'* patterns, the information encapsulated in the complement clause often carries traces of the writer's epistemic or attitudinal stance towards a given proposition, depending on the semantic properties of the noun controlling the clause (see Biber et al. 1999, p. 645). Classifying the nouns frequently used in these constructions based on their meanings, Charles (2007) found that writers in Politics primarily used argument nouns (e.g., *argument, assertion*) to take a stance towards the work of others. In comparison, writers in Materials science showed a tendency to employ these lexico-grammatical patterns with evidence nouns (e.g., *evidence, observation*) to evaluate their own research. Based on these findings and also those of Charles (2006), Charles (2007) concluded that the construction of stance appears to be fundamentally different across natural and social sciences but that more research (based on a larger corpus and a wider range of disciplines) is needed to identify the lexico-grammatical realizations of this discourse strategy and shed further light on disciplinary specificity in stance taking.

While the research summarized above has focused on a single lexico-grammatical pattern, other studies have also concentrated on a collection of lexical and grammatical devices that fulfill a particular communicative function. For instance, Hyland (1998) investigated various lexico-grammatical markers of 'hedging' (e.g., *may possibly be due to*, *that we consider most probable*) and 'boosting' (e.g., *it is well-known that, this is the central point*). Both categories of these lexico-grammatical features help writers modify their statements and knowledge claims by intervening into their discourse and engaging with the reader (see Hyland, 2000; 2005). Drawing on a corpus of 56 research articles from across

eight disciplines, Hyland (1998) observed that the use of hedging devices was more frequent than that of boosters, suggesting that writers in general tend to rely more on 'mitigation' than 'emphasis' when modifying their knowledge claims. Hyland's (1998) analysis, however, also revealed a wide disciplinary variability in the use of these devices, with over 70 percent of all hedging devices occurring in Humanities and Social sciences (Philosophy, Marketing, Linguistics). Hard knowledge fields such as Physics and Engineering were also underrepresented in the number boosters compared to Philosophy and Marketing. Hyland's (1998) results also revealed that, in general, research articles written in soft knowledge fields contained about two and a half times as many hedging and boosting devices as did hard sciences research papers. Hyland's (1998) findings are important as they indicate that the linguistic choices made by authors to fulfill a particular communicative purpose in research writing are often firmly anchored in conventional disciplinary norms and practices.

More recently, Afros and Schryer (2009) investigated the role of disciplinarity in the linguistic realization of another common communicative function in academic research writing: self promotion. For this purpose, the authors examined a collection of lexico-grammatical features often associated with this communicative function in research writing: personal pronouns (*I, we*), evaluative lexis (*it is evident, it is desirable to*), coordination (*and, but*), lexical cohesion (synonymy, antonymy, repetition) and comment clauses (*as I have*). Using a corpus of 20 single-authored research articles written in Linguistics and Literary studies, Afros and Schryer (2009) found both similarities and distinctions in the distribution of the examined features across the two disciplines. In particular, it was observed that both literary and linguistics scholars tend to rely on evaluation lexis in the Introductions and Discussions of their papers. However, Afros and Schryer (2009) also observed that the evaluation lexis used in Discussions in Literary studies was often accompanied by intensifiers (*deliberately, clearly*), while the use of evaluation lexis in the same section in Linguistics

often co-occurred with self-citations and first person pronouns. Afros and Schryer's (2009) observations indicate that specific discourse functions can give rise to consistent co-occurrence patterns of certain linguistic features and that these patterns can substantially vary, depending on different practices of disciplinary communities. In other words, the co-occurrence of linguistic features in research writing is highly unlikely to be coincidental; rather, it is systematically influenced by specific discursive functions and particular situational parameters, such as disciplinary conventions (more on this in Chapter 6).

The above reviewed research has used a deductive (top-down) research approach in which the theoretical framework for investigating lexico-grammatical patterns is determined at the outset of the research: the communicative function of interest is selected prior to corpus investigation, and the entire analysis is then conducted on this a priori basis (see Section 2.3.4.2; Biber, Conrad, Upton, 2007 for further discussion). The opposite approach would be to use quantitative methods to locate (through statistical analysis) patterns of linguistic features that systematically co-occur and then identify the underlying discourse functions that give rise to their co-occurrence. In this bottom-up analytical approach, which was developed by Biber (1988), communicative functions underlying co-occurrence patterns of linguistic features emerge from the corpus analysis with no preconceived functional/discourse theory guiding their identification. The development of this analytical technique, referred to as 'Multi-Dimensional' (MD) analysis, was motivated by the premise that the interaction between various means of language and their underlying discourse functions can highlight various purposes of human communication (see Biber, 1988, p. 13). Biber (1988) used this research method to identify the functional dimensions that relate to the co-occurrence patterns of lexico-grammatical features across a range of texts (e.g., transcripts of face-to-face conversations, personal letters, official documents, academic prose; extracted from two publicly available corpora: Lancaster-Oslo-Bergen Corpus of British English and the

London-Lund Corpus of Spoken English). In doing so, a wide range of lexical and grammatical features (67 linguistic devices) were first identified and 'tagged' in the selected texts, using a computer program developed by Biber (1988). Another specialized computer program was then developed by Biber to count and calculate the normed rates of occurrence for the tagged features. In the third step, Biber (1988) used the multivariate statistical method of factor analysis to determine significant co-occurrence relations among the features. In the final step, the quantitative data (factors) provided by the factor analysis were interpreted to determine the functions underlying the identified statistical co-occurrence relations (see Biber & Conrad, 2009 and Chapter 6 of the present dissertation for a detailed discussion of these steps).

Biber's (1988) analysis resulted in the identification of six functional dimensions. It is important to note that Biber's (1988) analysis resulted in seven factors, six of which were interpreted and proposed as functional dimensions. Each of the six dimensions provided an overall characterization of different spoken and written registers. For instance, Biber (1988) showed that academic writing has the characteristics of informational production (Dimension 1) and is highly abstract in nature (Dimension 5); whereas face-to-face conversation is highly involved, interactive and affective (Dimension 1) and presents non-abstract information (Dimension 5) (see Biber, 1988, Chapter 7 for a detailed discussion). Such functional interpretations were based on close examinations of the co-occurring features in texts with highest scores on each dimension. For example, for the first dimension, Biber (1988) observed that WH questions (e.g., *what did you do?*), WH clauses (e.g., *I believed what you told me*), first and second person pronoun (e.g., *I, you*), that-deletion (e.g., *I think [that] I'll go*), contractions (e.g., *it's* been), causative subordination (e.g., *because*), discourse particles (e.g., *anyhow*), private verbs (e.g., *think*) co-occurred more frequently in telephone or face-to-face conversations than in official documents or academic prose. Focusing on the pragmatic

and practical meanings conveyed by the co-occurrence of these features, Biber (1988) observed that WH questions, first and second person pronouns, private verbs in present tense form often co-occur in primarily interactive discourse. In addition, Biber (1988) observed that academic prose were characterized by the use of long words (word length), precise vocabulary (high type/token ratio), nouns, attributive adjectives that further elaborated nominal information, and prepositional phrases serving to integrate high amount of information into a text. Biber (1988) concluded that the high frequency of these features in academic prose can be associated with the high informational focus of these texts, which requires a careful integration of information in discourse. Biber's (1988) findings provide a clear indication of how a bottom-up analysis of a wide range lexico-grammatical features can highlight multiple dimensions of variation among texts.

In the past three decades, a multitude of studies have used Biber's (1988) MD analysis to examine variation in co-occurrence patterns of lexico-grammatical features among spoken and written registers. Research in this area has adopted two approaches for applying MD analysis to the study of linguistic variation in texts. In the first approach, referred to as 'additive' MD analysis, researchers apply previously defined functional dimensions (often those identified in Biber, 1988) to texts in their corpus. An additive MD analysis does not reveal new functional dimensions; instead, it complements the dimensions identified in previous MD studies. Conducting an additive MD analysis is straightforward as it basically involves calculating dimension scores for texts in the corpus under analysis based on the mean and standard deviation scores reported in the reference study (for further discussion see Sardinha, Pinto, Mayer, Zuppardi, & Kauffmann, 2019). This approach is mainly top-down, where variation among texts is posited on an a priori basis (i.e., the existing functional dimensions). In the second approach, commonly known as 'full' MD analysis, factor extraction procedures are carried out to arrive at 'new' functional dimensions specific to the

corpus under investigation (see Biber & Conrad, 2009; Sardinha & Pinto, 2019 for a detailed discussion of these two different approaches). There are a number of steps that need to be taken in order to conduct a full MD analysis of the data (see Egbert & Staples, 2019; Friginal & Hardy, 2019). At each step, there are also several key decisions that need to be made by the research so as to accurately identify co-occurring linguistic patterns and define their underlying functional dimensions (see Chapter 6).

Both types of MD analyses can provide important insights into linguistic variation among texts. However, a full MD analysis is technically more demanding than an additive MD analysis, as it requires an adequate knowledge of multivariate statistical techniques, such as factor analysis. In addition, because of its inductive and evidence-driven nature, a full MD analysis has the potential to lead the researcher to hitherto unknown dimensions of linguistic variation. It should be noted that the MD analysis used in Biber (1988) is a full MD analysis, which is mainly bottom-up and assumes no theoretical presumptions at the outset of the research.

Adopting these approaches, studies have applied MD analysis to the investigation of linguistic variation within highly specialized registers, such as academic (research) writing. For instance, Conrad's (1996b) dissertation research analyzed linguistic patterns of variation in academic writing across two disciplines: Ecology and History. In doing so, she conducted an additive MD analysis, in which texts in her corpus were mapped onto Biber's (1988) dimensions. Conrad's (1996b) results revealed disciplinary variation in academic research writing in Ecology and History. For instance, it was found that Ecology research articles scored less than History papers on Biber's (1988) second dimension (narrative versus non-narrative concerns), suggesting that research writing in Ecology, compared to History, is less narrative in nature. In a similar study, Biber and Finegan (2001) adopted an additive MD approach to examine linguistic variation across the main four sections (IMRD) of Medical

research articles. Biber and Finegan's (2001) analysis showed that research writing across all sections of Medical research articles was essentially informational and non-narrative (based on dimensions 1 and 2 identified by Biber, 1988). However, it was also observed that the discourse of methods sections was more impersonal (dimension 5) than those of introduction, results and discussion sections.

More recently, Gray (2015) conducted a full MD analysis of research writing across six disciplines: Philosophy, History, Political science, Applied Linguistics, Biology, and Physics. Gray (2015) classified articles written in the six fields based on their situational characteristics into theoretical (e.g., Philosophy), qualitative (e.g., Applied Linguistics), and quantitative (e.g., Physics). Gray's MD analysis of these registers identified four functional dimensions: academic involvement and elaboration versus information density (Dimension 1), contextualized narration vs. procedural description (Dimension 2), human vs. non-human focus (Dimension 3), academese (Dimension 4). The first dimension revealed that the discourse of Philosophy articles is highly involved and elaborate, marked by the co-occurrence of lexio-grammatical features such as conjunctions (e.g., *if, and, nonetheless*), finite clauses (e.g., *that*-clauses controlled by verbs, nouns, and adjectives), pronouns (e.g., first person pronouns), evaluative attributive adjectives (e.g., *important, simple*), among others. The same dimension also revealed that quantitative Biology articles were characterized by the co-occurrence of process nouns (e.g., *action, activity*), past tense verbs, agentless passives (e.g., *rates are used to refer*), passives as postnominal modifiers (non-finite -ed clause postmodifying a nouns, *phenological events divided by soil temperature change*), prepositions, long words, lexical precision (high type/token ratio). The co-occurrence of these features was taken as a reflection of the highly compressed and informational styles of discourse in such disciplines as Biology. As is often the case in MD studies, Gray's Dimension 1 included the greatest number of linguistic features, in

comparison with the other identified dimensions (also see Biber, 1988, p. 104). However, the other three dimensions also revealed interesting results. For example, dimension two showed that Physics articles are often characterized by procedural descriptions, whereas writing in qualitative History, Politics, and Applied Linguistics papers is often focused on contextualizing claims and interpretations within a narrative that describes events as evidence. Gray's third dimension demonstrated that the discourse of qualitative Applied Linguistics papers is more human focused than that of quantitative Biology papers. The fourth and the final dimension, while containing the fewest number of features (also see Biber 1988, p. 114), revealed interesting variation in academic writing across disciplines, suggesting that there is a fundamental difference between the type of academic writing which is often viewed as 'empirical' in nature (e.g., quantitative Political science) and that which is less concerned with being promoted as explicitly empirical (e.g. qualitative History). Gray's (2015) observations, together with those of Conrad (1996b) and Biber and Finegan (2001), suggest that MD analyses of disciplinary writing can provide important insights into how conventions of academic fields can encourage writers to make specific linguistic choices, which in turn give rise to systematic co-occurrence patterns of lexio-grammatical devices.

2.4. Conclusion

The present chapter has sought to show that the modern academic system is organized around the different forms of knowledge that scholarly communities produce as a result of their research activities, that academia is not a unitary and undifferentiated mass, and that specialization plays an important role in the production of knowledge in academe.

The dissemination of scholarly knowledge through the medium of (written) language can also be differentially affected by the epistemological orientations, beliefs, and ethos of disciplinary communities and their routine scholarly practices. The present chapter has

attempted to explain that disciplinarity can give rise to systematic patterns of language use in academic research writing and that identifying such patterns at different levels of linguistic granularity can provide important insights into how the transmission of scholarly knowledge is carried out across academic disciplines. Such an investigation can afford the opportunity to establish a theoretical basis for a conceptual model of research writing across disciplinary fields.

Chapter 3

Materials and methods

3.1. Introduction

The focus of the present chapter is on the methodological procedures undertaken to build the

Disciplinary Corpus of Research Articles (*DCRA*), the source of data used for the analyses

which were carried out to address the research questions (see Section 1.3). The chapter

begins with a detailed explanation of the criteria and operational definitions used for

determining the corpus design. It then describes the sampling procedures undertaken to

compile the *DCRA* as well as how text files were made ready for inclusion in the corpus.

Finally, the chapter presents the makeup of the *DCRA* along with an overview of the different

linguistic analyses that were carried out on the corpus.

3.2. *DCRA*: corpus design

In order to arrive at an optimal design for a corpus that is intended to be representative of a

particular register (e.g., expert research writing), it is necessary to specify the characteristics

of the target texts that are to be included in the corpus (Biber & Conrad, 2009, p. 31; Biber &

Jones, 2009). These characteristics include factors such as mode (speech, writing, singing),

communicative purpose (general, specific), setting (place of communication, time), and

production circumstances (scripted, real time) (Biber & Conrad, 2009, Chapter: 2; Conrad,

1996; Gray, 2015, Chapter 4). Biber and Conrad (2009) refer to such parameters as the

situational characteristics of texts that provide important information about their non-

linguistic aspects and help characterize the register they represent. These characteristics can,

of course, vary depending on the text category under consideration. For example, academic

journal articles, as a text category most representative of expert research writing (e.g.,

Hyland, 2016), are often characterized by situational parameters such as the journals in which they are published, the disciplines to which they belong, their type (theoretical or empirical), their internal structure, and the status of their author(s) and their target readership. Such factors are non-linguistic characteristics of academic journal articles which can potentially affect the ways in which language is used in this genre of writing (Biber, 1994; Conrad, 1996b). Therefore, identification of such non-linguistic factors is a fundamentally important consideration for designing a corpus that is intended to represent linguistic variation in a particular context of language use. In what follows below, I describe the procedures undertaken to systematically include this consideration in the process of designing the *DCRA*.

3.2.1. Disciplines and article types

The first step in the identification of the non-linguistic characteristics of articles was to develop a list of well-established disciplines to be included in the *DCRA*. The disciplines were selected according to their epistemological orientation toward either the soft sciences (e.g., applied linguistics, management, sociology) or the hard sciences (biology, physics, chemistry). The disciplinary orientation of each field was determined with reference to the distinction between soft and hard disciplines characterized by Biglan (1973a), Hyland (2004), and Durrant (2017). Using this organizing principle, a list of 12 disciplines was created, comprising biology, business, chemistry, dentistry, management, mathematics, mechanical engineering, applied linguistics, philosophy, physics, politics, and sociology.

Following this, an inductive survey of high-ranking journals (based on their 5-year impact factor published by Thomson Reuter's *Web of Knowledge ISI*) was conducted to identify the type of articles commonly published in the selected disciplinary fields. It is important to note that, in keeping with the aims of the present research (described in Chapter 1), the focus of this survey was on full-length articles and, therefore, short pieces such as book reviews,

commentaries, notes and responses were excluded from further examination. The journal survey process involved reading carefully journal descriptions and examining the content of the articles published in each volume/issue. Great care was taken to survey both general and more specialized journals within each discipline. The rationale for this decision was that certain disciplines, such as physics, include sub-areas which are theoretical in nature (e.g., theoretical physics). Following this initial survey, articles in at least 10 volumes of each journal were carefully examined based on the following characteristics to determine their type:

a) The inclusion of quantitative/qualitative data analyses

b) Presenting a new model or formula that is analyzed and tested through various analytical procedures

c) Focusing on theoretical aspects of the field without the inclusion of data

Articles belonging to (a) and (b) were grouped under 'empirical articles' and those identified as belonging to the (c) category were classified as 'theoretical articles'. It should be noted that, in cases where the researcher was in doubt regarding the type of an article, the article in question was discarded. For instance, in certain disciplines (e.g., physics, chemistry), writers were found to make methodological and theoretical arguments regarding the validity or accuracy of an established formula (or model) and present various data analyses to prove their point. Since such articles appear to have characteristics of both theoretical and empirical studies, they do not clearly fit into either category and, therefore, were not considered for inclusion in the corpus. Following the identification of article types, an additional survey was conducted to ascertain the extent to which journals in the selected disciplines publish each article type. Disciplines were then classified based on the type of articles they commonly produce. This classification is presented in Table 3.1.

Table 3.1. *Article types by discipline*

	Discipline	Empirical	Theoretical
Hard	Biology	++	+-
	Chemistry	++	+-
	Dentistry	++	-
	Mathematics	+	+
	Mechanical engineering	++	+-
	Physics	++	++
Soft	Applied linguistics	++	+
	Business	+	+
	Management	++	+-
	Philosophy	-	++
	Politics	+	++
	Sociology	++	+-

++very frequently occurs +commonly occurs +- rarely occurs - very rarely occurs (or not found)

The above classification provides an initial sense of variation in disciplines by shedding light on their theoretical and/or empirical orientations. For instance, as can be seen from the table, empirical research and data-supported studies are generally not present in philosophy. In line with the aims of the present study, empirical articles were selected as the target text type for linguistic analyses in this research -- these articles were also found to commonly occur across the majority of disciplines (Table 3.1). Consequently, articles published by Philosophy journals were not included in the corpus.

3.2.2. Textual layout and organization

The second step in the identification of the situational characteristics of articles was to analyze their internal structure. To this end, empirical papers in each discipline were examined to determine their organization structures. This analysis revealed two main organizational patterns: a) papers which included the content of the standard four-part organization (Introduction, Methods, Results, Discussion; IMRD) and b) articles which contained a different and rather distinct sectioning format. Empirical papers in most disciplines were found to belong to the former category, except for mathematics papers which contained distinct organizations. For consistency and comparison purposes, articles

from this discipline were excluded from further consideration, resulting in a total of 10 disciplines to be included in the corpus.

IMRD sections were identified based on section labels used by authors. That is, we did not rely on our own judgments for the identification of the sections. Instead, we relied on the labels that authors deemed appropriate for the description of the content of each section of their articles. Thus, for example, the Introduction sections included in the soft sub-section of the corpus contained the type of content that authors in the soft fields deemed representative of Introductions in these fields (the same holds for all IMRD sections included in the corpus). Literature Reviews/background sections were not considered for inclusion in the corpus. This decision was based on the observation that most research articles in the sciences did not contain literature review sections (cf. Lin & Evans, 2012; see Stoller & Robinson, 2013). In addition, the analysis revealed that the Results and Discussion (R&D) sections in empirical papers were either fully integrated or separated (cf. Stoller & Robinson, 2013). Every effort was made to ensure that all the articles included in the corpus contained fully separated R&D sections. The rationale for this decision was that R&D sections can perform distinct discourse functions and fulfill different communicative purposes in the RA (e.g., Lin and Evans, 2012; Swales, 2004). The difference in the communicative aims of R&D sections can in turn give rise to systematic patterns of language use that can vary as a result of the sections' communicative purposes. Therefore, it is important to maintain the distinction between these two sections and recognize the significance of their rather distinct communicative aims in discourse. In line with this reasoning, Conclusion sections were not treated as part of Discussions and those articles that contained integrated Discussion and Conclusion sections were not included in the corpus.

3.2.3. Writers and their language backgrounds

The third and final step in the identification of the situational characteristics of articles was to verify the status of their authors as either native or non-native speakers of English. To distinguish between L1 written articles and those authored by LX writers, the method for authorship categorization proposed by Wood (2001) was used as a basis for the sampling procedure. Wood's (2001) method is based on two criteria: broad and strict. While the 'broad' criterion takes the name of the first author to distinguish between L1 and LX written texts, the 'strict' criterion takes a step further by considering authors' institutional affiliations as an additional differentiating factor. Although this sampling method has been widely used in many studies (e.g., Hyland, 2016; Martinez, 2018; Peacock, 2002; Wood, 2001; Pen, Reppen, Biber, 2016), it is not entirely foolproof. For the present study, a conservative approach was taken, systematically implementing the following steps to optimize the process of distinguishing between L1 and LX written article. Specifically:

a) Each author's first and last names were examined

b) Each author's institutional affiliation was examined and cross-checked against current correspondence address

c) If (a) and (b) revealed that the article was multi-nationally authored, it was <u>not</u> considered for inclusion in the corpus

d) If (a) and (b) confirmed that the article was <u>not</u> multi-nationally authored, the researcher emailed the corresponding author of the article requesting clarification on the language background of the authors involved in drafting the manuscript

e) If no response was received from the corresponding author with regards to the inquiry, the research article in question was <u>excluded</u> from inclusion in the corpus

f) If the corresponding author confirmed that all the authors whose names were listed in the article were <u>native</u> speakers of English, the article in question was considered for inclusion in the corpus as a L1 written article. Similarly, if the corresponding author confirmed that all the authors were <u>non-native</u> English speakers, the article was considered as a LX written article.

Figure 3.1 and 3.2 provide examples of the articles that were included in the *DCRA* as L1 and LX written papers. On the contrary, Figure 3.3 shows an example of multi-nationally authored articles which were excluded from further consideration.

Research Article
Predicting protein subcellular localization based on information content of gene ontology terms

Shu-Bo Zhang [a] ✉, Qiang-Rong Tang [b] ✉

[a] Department of Computer Science, Guangzhou Maritime Institute, Room 803 Building 88, Dashabei Road, Huangpu District, 510275, Guangzhou, PR China
[b] Department of Shipping, Guangzhou Marine Institute, Room 205 Shipping Building, Hongshan NO.3 Road, Huangpu District, 510275, Guangzhou, PR China

Figure 3.1. Example of a LX article (Biology)

Cohesive finite element modeling of the delamination of HTPB binder and HMX crystals under tensile loading

David J. Walters [a], Darby J. Luscher [a] ✉, John D. Yeager [b], Brian M. Patterson [c]

[a] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
[b] Explosive Science and Shock Physics Division, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
[c] Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Figure 3.2. Example of a L1 article (Mechanical engineering)

Recent Developments in Functional Crystals in China

Jiyang Wang [a] ✉, Haohai Yu [a], Yicheng Wu [b], Robert Boughton [c]

[a] State Key Laboratory of Crystal Materials, Shandong University, Jinan 250100, China
[b] Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100080, China
[c] Department of Physics and Astronomy, Bowling Green State University, Bowling Green, OH 43403-0001, USA

Figure 3.3. Example of a multi-nationally authored article <u>not</u> considered as a LX written RA

3.3. Corpus compilation, file conversion and text clean-up

For the ten remaining disciplines, research articles were extracted from high-ranking journals (as reflected by their Thomson ISI 5-year journal impact factor). The extraction of articles was based on a stratified random sampling procedure. Five articles were sampled from different volumes of each journal published between 2000 and 2018. This was done to preclude the possibility of data skewed by oversampling from one particular volume or issue. Each text was then thoroughly examined to ensure that it followed the operational definitions and criteria for article selection described in sections 3.2. That is, the extracted text was selected for inclusion in the corpus only if it matched the operational definitions for determining article type (i.e., empirical research papers), organizational structure (i.e., the IMRD framework), and the language status of authors. If it did not, then a different article was selected from the same volume. This sampling procedure was repeated until the target number of texts was reached for both L1 and LX corpora.

Following the sampling procedure, all the sampled files were converted to plain text (.txt). The converted files were then manually edited and cleaned out of any marks that were not part of the prose of the article, including page headers, page numbers, superscripted numbers, special symbols, and formulas. All tables, figures and reference lists were also removed from the sampled files. Section markers (i.e., headings) were retained and annotated. Extra markup information was also added to each file, specifying particular features of the article such as its title, the journal and the discipline from which it was sampled, and the name(s) of its author(s) and their language background(s). These indicators were enclosed in angled brackets and were placed at beginning of each file. The IMRD sections were then identified (through their corresponding section markers) in the articles and stored separately for both L1 and LX corpora.

3.4. Corpus description

The *DCRA* is composed of 600 research articles (c. 4 million words) written by L1 and LX authors in 120 academic journals, representing 10 different disciplines. L1 and LX corpora of the *DCRA* are closely matched both for the total number of words (c. 2 million words in each) and for the number of texts (300 in each). Table 3.2 provides a description of the *DCRA* and its subcorpora.

Table 3.2. *Composition of the DCRA*

Discipline	L1 corpus	LX corpus
Biology	245878	251508
Chemistry	214974	240812
Dentistry	211645	212714
Mechanical engineering	206385	198070
Physics	197368	186865
Hard total	**1076250**	**1089969**
Applied linguistics	210760	203337
Business	258268	246967
Management	229049	244412
Politics	221233	204351
Sociology	259833	224822
Soft total	**1179143**	**1123889**
Corpora total	**2255393**	**2213858**
Corpus total	**4469251**	

3.5. Overview: Corpus analysis and quantitative methods

Most of the corpus analyses in this research were carried out using publicly available software tools and specialized computer programs developed by the researcher. These analyses included corpus searching, automatic tagging, concordance building, and various quantitative data analyses. The quantitative procedures were also complemented by qualitative discourse analyses in order to provide further contextualization of the patterns observed in the corpus and describe the obtained data in a wider context.

These corpus and data analyses are explained in more detail in the subsequent chapters (Chapters 4-6). These chapters present a series of studies that use different methodologies to investigate the distribution and use of different linguistic variants in the

DCRA. The first of these studies, which is presented in Chapter 4, uses Gries' (2008) Deviation of Proportions statistics, keyword analysis, and hierarchical clustering to investigate the use of single words in the corpus. Study 2 (Chapter 5) takes as its unit of analysis a particular kind of multi-word expressions, namely lexical bundles, and uses inferential statistics to make inferences about the use of these items in research writing (as presented in the *DCRA*). Study 3 (Chapter 6) uses multi-dimension analysis to investigate key dimensions of linguistic variation as reflected in the co-occurrence of a wide range of lexico-grammatical features (e.g., appositive noun phrases, *that*-complement clauses) in the corpus. The remainder of the thesis provides a more detailed explanation of the procedures undertaken to carry out the analyses in each of these three studies and discusses the significance of their findings for identifying the unique characteristics of research writing across disciplines.

Chapter 4

Study 1: Parameters of variation in the use of single words in academic research writing

4.1. Introduction

Writing plays a central role in the creation and dissemination of knowledge in academia (see Russell, 1991). Due to its distinctive style and conventions, academic writing has spawned major investigatory efforts in both research and practice across various fields of inquiry such as philosophy, history, sociology of knowledge, rhetoric, language education, and applied linguistics (see Chapter 2, Section 2.2). Research in the applied areas of this interdisciplinary inquiry has primarily focused on creating pathways to achieving proficiency in academic writing. One promising avenue that has long been a topic of interest in applied studies and educational circles in this area is academic vocabulary knowledge. Research has shown that ample knowledge of such vocabulary can be conducive to gaining academic literacy in that it provides students with varied lexical choices that are particularly useful for communication in academic discourse (Corson, 1997; Coxhead & Nation, 2001; Li & Schmitt, 2009; Snow & Kim, 2007).

This vocabulary is typically classified into two sub-categories: sub-technical and technical vocabulary (Hierber & Lubliner, 2008; Nation, 2001). Sub-technical (or general academic) vocabulary refers to those items that are not tied to a specific subject area and are commonly used in a wide range of academic disciplines (Coxhead, 2000). This vocabulary, which includes lexes such as *insight*, *exhibit*, *collapse, adequate,* falls somewhere between non-academic and technical vocabulary (Beck, McKeown, & Kucan, 2002, 2008; also see Coxhead, 2019). Technical (or discipline-specific) vocabulary is a set of subject-related items (e.g., *estrogen, periodontal, nucleation*) that are used to create specialized knowledge in a

given domain of inquiry (Chung & Nation, 2004). Such domain-specific lexes are employed by members of a particular academic community to construct a specialized form of discourse which is exclusive to their target audience and may not be readily understood by members of other disciplinary domains.

Knowledge of academic vocabulary (both technical and sub-technical) plays an important role in the construction of meaning in written academic genres. A number of studies have explored the specific use of academic vocabulary in disciplinary fields (Durrant, 2014, 2016; Hyland & Tse, 2007; Konstantakis, 2007; Martinez, et. al., 2009). The rationale for investigating disciplinary specificity in the use of academic vocabulary is two-fold: (a) some words are highly likely to be more important for academic (research) writing in certain disciplines, compared to others (e.g., Durrant, 2016; Hyland & Tse, 2007; Martinez, Beck, and Panza, 2009) and (b) it is highly unlikely that these words are evenly dispersed across disciplines, and thus they may not be equally useful for writing in all branches of academic study (e.g., Durrant, 2014). In line with this argument, several studies have investigated disciplinary variation in the use of items from academic vocabulary lists (e.g., the AWL and the AVL) in a range of academic disciplines (e.g., Durrant, 2016; Hyland & Tse, 2007; Martinez, et al, 2009). The evidence from this line of research appears to endorse the notion of specificity in academic vocabulary use. However, since these studies mainly focus on the evaluation of such specificity on the basis of previously compiled academic wordlists, it can be argued that their conclusions are largely restricted to the items included in those lists. The approach adopted in these studies is to use corpus evidence to analyze different patterns of use for a set of pre-defined items. In other words, corpus analysis in such studies is viewed as an evidential tool to validate theoretical presumptions about a collection of linguistic items (see, for example, Siyanova-Chanturia & Omidian, 2020 for a discussion). However, more generalizable conclusions can be reached by adopting an inductive approach in which

patterns of disciplinary specificity in vocabulary use emerge from the analysis of a corpus with minimal a priori assumptions guiding their identification. Such an approach has the potential to shed light on the steps that should be taken to explore various aspects of the observations arising from the data (Baker & Egbert, 2016; Partington, Duguid, & Taylor, 2013; more on this in section 4.3). In this approach, it is the emerged linguistic patterns that lead the researcher to a linguistic theory, and not vice of versa.

Very few studies to date have adopted such an approach to investigating disciplinarity in the use of single-word items (Durrant, 2014). Fewer still have adopted this approach to examine variation in the use of such vocabulary in academic research writing and across different parts of research articles in various disciplinary fields. As was mentioned earlier (see Chapters 1 and 2), one of the primary means of disseminating knowledge in academic contexts is research writing. Academic research writing is regarded as a highly conventionalized genre in which effective communication of knowledge is contingent on writers' familiarity with the linguistic expectations of their target readership (Bhatia, 1993; Hyland, 2001; Swales, 1990). As Hyland (2011) argues, in writing high-stakes genres such as research articles, in which effective communication of meaning involves the anticipation of alternative interpretations and readers' possible objections, authors often tend to rely on linguistic resources that can reflect the disciplinary values and expectations of their readers. Such linguistic expectations are mostly related to the use of words (or word sequences) that are conventionally expected by readers in a given academic community.

Such considerations, together with the fact that research writing is now an integral part of academic life across disciplines (Hyland, 2016, Flowerdew, 2013; also see Section 1.2), stress the importance of providing a comprehensive picture of the similarities and differences in the language of research writing across disciplines. As was noted above, focusing on a set of pre-defined items (such as those included in vocabulary lists) for

highlighting these commonalties and variations can curtail the generalizability of the patterns that emerge from such an investigation. For this purpose, the present study set out to investigate parameters of variation in research writing across disciplines by adopting an inductive approach in which patterns of specificity in vocabulary use emerged from corpus data with no preconceived assumptions. The following section describes the methodology used in the present study in more detail.

4.2. Methodology

To quantify linguistic variation in research writing in terms of word use, two complementary measures were used: lexical dispersion and keyness. The former measure was employed to determine the degree to which frequently used vocabulary in research writing is evenly dispersed across academic disciplines and different sections of research articles. This measure allowed us to classify high-frequency words in research writing based on their level of technicality without relying on any pre-defined categorizations (e.g., technical and sub-technical vocabulary; see Nation, 2001). Variation in research writing across disciplines and IMRD sections was then characterized based on this classification. Following this, the measure of keyness was employed to verify and further explore the patterns identified by the lexical dispersion measure. For this purpose, commonality and variation between disciplinary fields and sections of research articles were determined based on the degree of overlap in key vocabulary use (more on this below). This analysis afforded a mapping of specificity in research writing on the basis of the vocabulary that plays a key role in the communication of knowledge across academic disciplines and IMRD sections. In what follows below, I provide a more detailed explanation of the steps taken to operationalize these two measures.

As a first step, separate listings of high-frequency words were first generated for the DCRA and different sub-sections of the corpus: disciplines and the IMRD sections in each

discipline. Words were considered frequent if they (a) occurred more than 100 times per million words and (b) appeared in at least 10% of texts in the corpus (see, for example, Durrant, 2014). It is also important to note that 'vocabulary use' in this study was operationalized as the use of individual *word forms* (as opposed to lemmas and word families) in research writing. The rationale for focusing on word forms as the target unit of analysis was to account for their instability in use across different discourse types. Corpus-based vocabulary studies have repeatedly shown that different forms of a word can exhibit disparity in the semantic, syntactic, and thematic ties that they establish with their surrounding context (e.g., see Durrant, 2014, 2016; Hyland, 2008; Hyland & Tse, 2007; Sinclair, 1991). This level of disparity in use is often influenced by the situational or non-linguistic aspects (i.e., disciplinary conventions) that characterize the register in which word forms are used (see Gardner, 2008 and Section 3.2. for further discussion). In the case of the present study, these situational characteristics pertain to disciplinary conventions and specific communicative purposes of different sections of research articles. For example, Hyland and Tse (2007) showed that the derivationally-related forms in the word family analyse on the AWL were used to varying degrees across disciplines, with the sciences writing making greater use of the adjective form analytical and the Social Sciences relying more on the noun form analysis. Previous research has also shown that certain inflectional forms are used more or less frequently across sections of research articles. For example, the past tense and past participle forms of verbs are typically more common in Methods as compared with Introductions and Discussions (e.g., see Biber & Conrad, 2009, p. 130; Martínez, et al. 2009; also see Gray, 2015, p. 104). Ignoring such variation across different forms of a word by collapsing them into more abstract categories (e.g., lemmas and word families) would, therefore, run the risk of obscuring the linguistic clues that can highlight possible variation in vocabulary use in academic research writing (also see Durrant, 2014).

Having created the lists of high-frequency words, I then used Gries's (2008) *deviation of proportions* statistic to assess the extent and nature of variability in vocabulary use in the corpus and its subsections. Deviation of proportions (henceforth DP) is a statistical measure that determines the extent to which a given word form is evenly dispersed across different parts of a corpus. It should be noted that corpus parts in the present study were operationalized on the basis of the situational parameters characterizing texts in the corpus (i.e., academic disciplines and IMRD sections, see Egbert, Burch, & Biber, 2020 for a detailed discussion of measuring lexical dispersion across linguistically meaningful corpus parts). To calculate DP, the following steps were taken in the present study (as suggested Gries, 2008, p. 415):

- Determining the size of each sub-section of the DCRA and normalizing it against the overall size of the corpus to arrive at <u>expected percentages</u> of word forms while taking into account the size of each section.

- Determining the frequency with which word forms occur in a given section and normalizing it against the overall number of their occurrences in the entire corpus, creating the <u>observed percentage</u> of use for each word form in each section.

- And finally, computing all pairwise 'absolute' differences between <u>expected</u> and <u>observed</u> occurrences of word forms across sections, summing these differences and dividing them by two.

To further examine the patterns of vocabulary use highlighted by DP, important words in each section of the corpus were identified using Scott's (1996) concept of *keywords*. According to Scott (1996), keywords are words which have a special status in a given text (or register). A single word is considered a *keyword* if it occurs more frequently in a corpus than would be predicted by chance when compared to a reference corpus (Baker, Hardie, &

McEnery, 2006; Paquot, 2007; Scott, 1997). The degree to which a word type is 'overused' in a target corpus compared to a comparison corpus determines its 'keyness' value, which is typically measured by chi-squared or log-likelihood ratio tests (see Scott & Tribble, 2006). It should be noted that, since log-likelihood tests are thought to provide more accurate results than chi-squared tests when comparing the rates of occurrence of rare events, such as keywords (see Dunning, 1993), log-likelihood tests were used to measure the keyness values of words in the present study (also see Brezina, 2018, Chapter 3 for further discussion).

Separate listings of keywords were generated for the DCRA and its subsections using WordSmith Tools (version 7.0, Scott, 1999), taking the British National Corpus (BNC)[2] as a reference corpus. The percentage of overlap between the generated keywords lists was then calculated in a pairwise manner to assess the degree of commonality in the use of important words between different subsections of the corpus (see Durrant, 2008, 2009 for applications of the same method; see Chapter 2, Section 2.3.4.1). This analysis resulted in a matrix of overlap which was then used as a basis of a hierarchical cluster analysis. Cluster analysis is an *unsupervised* statistical method whose goal is to ascertain whether observations fall into relatively distinct groups without taking any a priori assumptions regarding their relations (see James, Witten, Hastie, & Tibshirani, 2013, p. 389). Hierarchical clustering can be classified into two main categories: hierarchical agglomerative clustering and divisive clustering (ibid). Hierarchical agglomerative clustering uses a bottom-up approach to build a hierarchy of clusters based on the homogeneity of data points in each group. In contrast, the less commonly-used divisive clustering is based on a top-down approach in which all data points are grouped under a single cluster and then heterogeneous clusters are separated in each iteration. Since the implications of the agglomerative clustering method proved more

---

[2] The BNC is a 100-million-word collection of over 4000 samples of English (see http://www.natcorp.ox.ac.uk/corpus/index.xml). The corpus has been widely used in linguistic research in the past two decades. The BNC has the advantage of being one of the largest English corpora freely available to researchers (see Hawtin, 2018 for further discussion).

practical for describing our data in terms of their quantitatively-defined commonalities, this technique was selected for the cluster analysis in this study.

4.3. Analysis and results

Once the lists of high-frequency words were retrieved from the DCRA, the first step in the analysis was to assess the extent to which frequent words in the corpus were evenly distributed across different disciplines. As was mentioned in the previous section, Gries' (2008) proposed dispersion measure (DP) was used for this purpose. DP scores for all frequent words in the corpus were calculated and were then normalized based on the DP normalization method proposed by Lijffijt and Gries (2012). The normalization procedure was carried out to preclude any chance of DP not reaching its maximal value (see Lijffijt and Gries, 2012). DP is maximal when all occurrences of the target word are found in the smallest part of the corpus under analysis (ibid). Factors such as the number of parts of the corpus or variation in their sizes have the potential to prevent DP from reaching its maximal value. Therefore, to guard against such potential disruptions, all DP values in this study were normalized using the following formula (adopted from Lijffijt & Gries, 2012):

$$DP_{norm}=DP/1-min(s)$$

where *min(s)* is the size of the smallest part of the corpus (e.g., the Dentistry sub-corpus in the DCRA) normalized against the overall corpus size.

The obtained DP_{norm} values for high-frequency words in the DCRA were found to vary substantially, ranging from 0.031 to 0.973. To aid interpretation, all words in the corpus were classified into four major categories based on their DPs: (1) extremely evenly distributed, with a DP_{norm} below 0.25, (2) moderately evenly distributed, with a DP_{norm} ranging from 0.25 to 0.499, (3) moderately unevenly distributed, with a DP_{norm} ranging from

0.5 to 0.749, and (4) extremely unevenly distributed, with a DP_{norm} value of 0.75 or above.

Table 4.1 presents the percentage of words falling within each dispersion range, along with a random selection of words from each. As can be seen from Table 4.1, the extreme point at the far left of the continuum represents instances of 'general' vocabulary (*for, of, and, however*), which accounted for 12.1% of all words in the corpus. These words clearly distinguish themselves from those residing at the opposite end of the continuum (i.e., extremely unevenly dispersed), which appear to be more technical in nature. These are specialized words (*entrepreneurial, specimen, precipitation, solar*) which seem to be key in the construction and transmission of field-specific meanings in research writing. These words were found to make up more than one-fifth of high-frequency words in the DCRA. Beyond these two extremities of the continuum, the level of specificity appears to also slightly vary between the words grouped under the second and third categories (i.e., moderately evenly dispersed and moderately unevenly dispersed, respectively). As we can see, those words that were found to be moderately unevenly distributed seem to be relatively more specific than those with moderately even distributions. As shown in Table 4.1, the former category (i.e., moderately unevenly dispersed) accommodates about two-fifth of words in the corpus. Classifying the four categories into two broad groups of 'more widely dispersed' (i.e., extremely and moderately evenly dispersed) and 'more narrowly dispersed' (i.e., extremely and moderately unevenly dispersed), we find that the latter group contains 1.6 times (61.5%) more items than the former (38.2%). This indicates the importance of this group of vocabulary in the production of scientific knowledge in research articles across academic disciplines.

Table 4.1. *Examples of high frequency words grouped under four different dispersion classifications*

	Cline of dispersion		
More widely dispersed		More narrowly dispersed	
Extremely evenly dispersed	Moderately evenly dispersed	Moderately unevenly dispersed	Extremely unevenly dispersed
0–0.249	0.25–0.499	0.5–0.749	0.75–1
828 (12.1%)	1791 (26.2%)	2698 (39.4%)	1513 (22.1%)
for; of; the; between; also; than; study; can; may; they; has; other; been; however; in; that; with; we; it; two; such; our; when; one; all; results; research; but; time; used; both; each; using; different; together; paper; taken; despite; yet high; level; effect; studies; only; based; social; there; first; effects; into; because; table; while; about;	events; obtained; complex; power; showed; age; features; network; method; function; strength; activities; respondents; items; initial; risk; states; experimental; interactions; focus; samples; status; characteristics; experience; context; activity; respectively; period; measure; quality; year; measures; relationship; values; observed; value; shown; positive; findings; support; participants; rate; range; hypothesis; growth; variable; structure; state; resources; behavior; theory; literature	disproportionate; constructions; innovation; undertake; stakeholders; pathway; synthesis; healthcare; governance; crises; modulation; monitored; ended; metallic; transactions; challenged; elicit; developers; controversy; negotiations; transverse; reputational; crossed; charges; casting; comparability; globalization; compensatory; comprehend; force; experiments; concentrations; containing; velocity; matrix; wave; sector; properties; organizational; conflict; complexity; maximum; ties; yield; curve; maintenance; compliance; commitment; zone; boundary	distilled; districts; longitude; propagating; subsidiary; republican; transnational; party; language; cells; teachers; strain; writing; learners; species; fracture; protein; news; binding; crack; entrepreneurs; entrepreneurial; load; precipitation; periodontal; specimen; solar; thickness; grain; electoral; deformation; acid; voting; earnings; politicians; toughness; immigration; gene; instruction; flux; campaign; war; democratic; radical; winter; neurons; military; resin; violence; ideological

To further examine these results, words grouped under these two broad dispersion categories (i.e., 'more widely dispersed' and 'more narrowly dispersed') were analyzed in terms of their degree of concentration in each discipline in the corpus. Table 4.2 presents the percentages of each category across disciplines.

Table 4.2. *Proportion of dispersion categories across disciplines*

	% Widely dispersed	% Narrowly dispersed
Biology	61.1	38.9
Chemistry	58.1	41.9
Dentistry	57.8	42.2
Mechanical Engineering	63.2	36.8
Physics	58.9	41.1
Mean	59.8	40.1
Minimum	57.8	36.8
Maximum	63.2	42.2
Applied linguistics	73.9	24.1
Business	73.3	24.7
Management	77.3	22.7
Politics	73.5	26.5
Sociology	75.7	24.3
Mean	74.4	24.4
Minimum	73.3	22.7
Maximum	77.3	26.5

As can be seen from Table 4.2, there are differences in the extent to which writers in different disciplinary fields draw on words from the two categories. On average, widely-dispersed words appear to outnumber narrowly-dispersed items across all disciplines. However, words grouped under the former category appear to account for only 60% of high-frequency word types in the hard sciences. Table 4.2 shows that two-fifth of individual words in research papers written in the hard disciplines were found to have a narrower and less widely distributional range. This means that, for every five individual words in these disciplines, two had a rather narrow distribution characteristic. In contrast, widely-dispersed words were found to have a greater concentration in the word lists retrieved from the soft sub-corpora, accounting for 74.4% of word types in these disciplines. Words with a narrower range in these disciplines, however, were found to have a mean percentage of 24.4%, which is about 1.65 times less than that in the hard sciences. This variation would suggest that research

writing in the hard fields demands the knowledge of a type of vocabulary that has a narrow range of use and applicability.

To further explore the observed disciplinary variation, a list of words that were found to be 'key' in each discipline was generated. These were words which (a) were used in the target discipline significantly more frequently than in the BNC, with the threshold for significance set at *p <.1E-7*, (b) appeared in at least 10% of texts in each discipline, (c) occurred at least 20 per million words in each discipline, and (d) contained at least two alphabetical characters (excluding acronyms). Following this, the percentage overlaps between the keyword lists of each discipline were computed. This procedure yielded a symmetric matrix representing the degree of overlap in the use of key vocabulary between all disciplines in the corpus. Table 4.3 provides more information about this matrix.

Table 4.3. *Overlaps in key vocabulary between disciplines in the DCRA*

	BI	CH	DE	ME	PH	AL	BU	MA	PO	SO	Mean	Max	Min
BI	---	51.1	40.6	36.8	36.1	26.5	29.6	31.1	27.9	30.6	34.5	51.1	26.5
CH	51.1	---	44.3	41.8	40.8	21.8	22.5	23.5	21.8	22.4	32.2	51.1	21.8
DE	40.6	44.3	---	36.6	31.2	23.6	25.4	25.9	24.5	25.5	30.9	44.3	23.6
ME	36.8	41.8	36.6	---	44.9	24.9	24.6	26.1	24.3	24.7	31.6	44.9	24.3
PH	36.1	40.8	31.2	44.9	---	21.2	24.6	24.4	25.7	27.5	30.7	44.9	21.2
AL	26.5	21.8	23.6	24.9	21.2	---	44.0	44.8	38.0	38.7	31.5	44.8	21.2
BU	29.6	22.5	25.4	24.6	24.6	44.0	---	66.7	49.0	50.6	37.4	66.7	22.5
MA	31.1	23.5	25.9	26.1	24.4	44.8	66.7	---	45.0	48.9	37.4	66.7	23.5
PO	27.9	21.8	24.5	24.3	25.7	38.0	49.0	45.0	---	50.9	34.1	50.9	21.8
SO	30.6	22.4	25.5	24.7	27.5	38.7	50.6	48.9	50.9	---	35.5	50.9	22.4

BI: Biology CH: Chemistry DE: Dentistry ME: Mechanical Engineering PH: Physics AL: Applied Linguistics
BU: Business MA: Management PO: Politics SO: Sociology

A number of points stand out from the matrix. First, the overlap values between disciplines appear to correspond to the *hard* and *soft* distinction proposed by Biglan (1973a), Hyland (2004), and Durrant (2017). As shown in Table 4.3, there is a low degree of overlap between hard and soft disciplines in terms of the key words they commonly use.

Secondly, the degree of homogeneity in the use of key vocabulary appears to vary even within the hard and soft disciplinary domains, with the confluence of certain disciplines

being stronger than others. For instance, there is a great deal of overlap between the key vocabulary used in biology and chemistry research papers. In fact, if we calculate the mean (40.42) and standard deviation (5.44) of overlap percentages for the cells belonging to the hard disciplines, we find that the percentage overlap between these two disciplines (i.e., chemistry and biology) is about two standard deviations above the average (z-score = 1.94), indicating their high degree of overlap compared to the other hard disciplines in the corpus. In addition, as can be seen from the matrix, mechanical engineering and physics also showed a greater degree of overlap with each other than with biology, chemistry, and dentistry. Among the soft disciplines, business and management were found to show a stronger overlap with each other (z-score =2.42) than they did with the other soft disciplines. And finally, applied linguistics was found to be an outlier among the soft disciplines in terms of key vocabulary use, as it showed the lowest average degree of overlap with these disciplines (M=41.3).

In order to extend this evaluation of these commonalities to a more quantitatively-defined level, a hierarchical cluster analysis was carried out based on the above matrix of overlaps between disciplines (see Durrant, 2008, 2014 for different applications of this approach). In doing so, Euclidean distance was used as the dissimilarity measure to determine the distances between disciplines based their degree of overlap (see James, et al, 2013, p. 402). Disciplines were then grouped according to the *average linkage* method which computes all pairwise distance dissimilarities between clusters and calculates the *average* of these dissimilarities to produce a hierarchy of the target variables (for a discussion of linkage methods in hierarchical clustering see James, et al, 2013, p. 399-402). This analysis produced the hierarchical arrangement of the disciplines based on the degree of overlap in their use of key vocabulary, as shown in Figure 4.1.
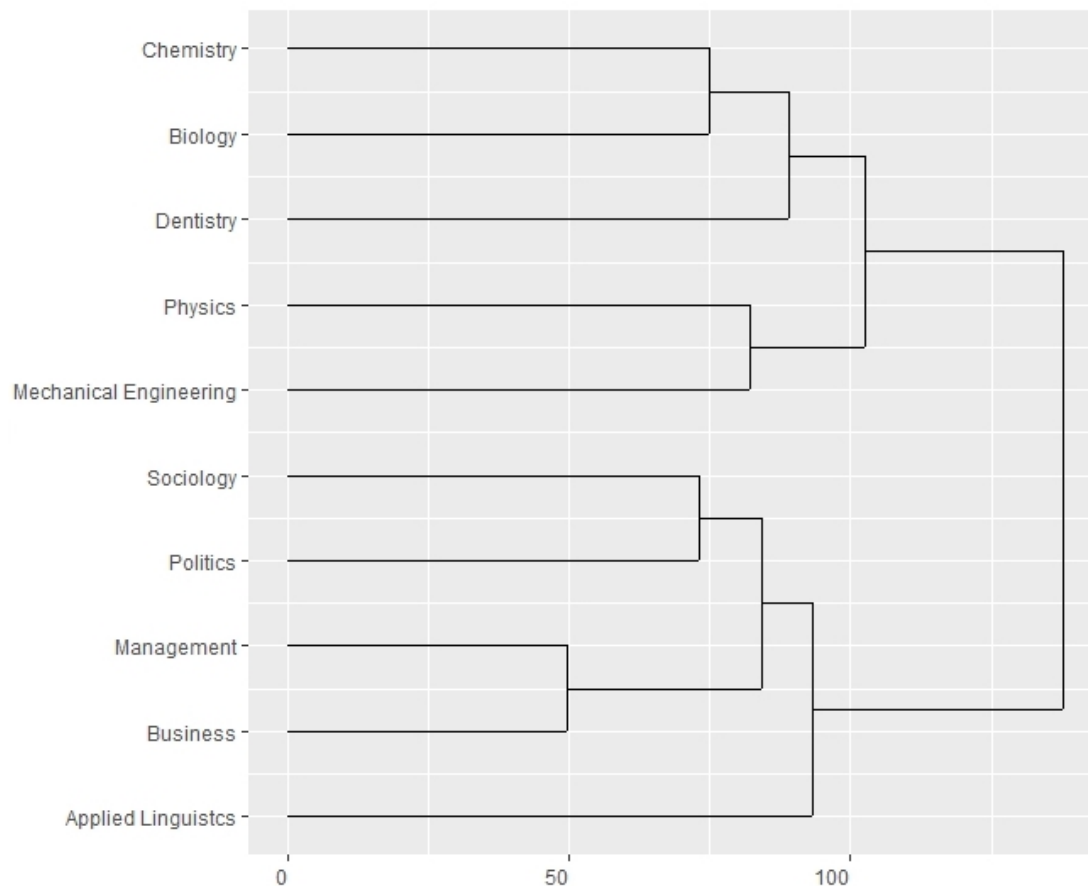
Figure 4.1: Cluster analysis of overlaps between disciplines

As shown in Figure 4.1, the cluster analysis distinguished the hard sciences (top half) from the soft fields (bottom half). Reading the top half of the dendrogram from left to right, we can see that biology and chemistry are assembled within the same cluster at the first level of the analysis. At the second level, physics and mechanical engineering form their own cluster. The biology-chemistry cluster is then joined by a single-point cluster formed by dentistry, which reflects a relative degree of homogeneity in the use of key words between these disciplines. This group of three disciplines then joins up with the mechanical-physics cluster to form a macro group of disciplines, representing the broad category of 'hard knowledge fields'.

Reading the bottom half of the clustering tree by the same logic, we find that management and business were the first to form a cluster due to their strong degree of overlap. In the next step, politics and sociology merge into their own cluster, which then joins

up with the management-business group. This group of four disciplines then combines with applied linguistics, creating the broad category of 'soft knowledge fields'. The fact that the single-point cluster created by applied linguistics joins the other four soft disciplines at the final steps of the analysis indicates its low degree of overlap with these disciplines and provides an indication of its outlier status as a soft field in terms of key vocabulary use. Altogether, these results suggest that the degree of homogeneity in the use of high frequency words that play a key role in the construction of knowledge in academic disciplines not only varies between hard and soft knowledge fields but also within these two broad disciplinary domains.

A more nuanced view of these differences can be achieved if we further explore variation in research writing practices in terms of vocabulary use across different sections of research articles. To achieve this, the DCRA was divided into four separate sub-corpora, each corresponding to the four main sections (i.e., IMRD) of the research article. High frequency words were then extracted from each sub-corpora based on the inclusion criteria mentioned in Section 4.2. This procedure was repeated for each section of research articles in each disciplines, resulting in 40 separate listings of high frequency words for the ten disciplines in the corpus. Gries's (2008) DP was once again used to assess the extent to which high frequency words were evenly distributed across different sections of research articles in different disciplines. Table 4.4 presents the degree of dispersion for frequent words in the four main sections of research articles across disciplines.

Table 4.4. *Degree of dispersion for frequent words in IMRD sections*

		More widely dispersed		More narrowly dispersed	
	Range	0–0.249	0.25–0.499	0.5–0.749	0.75–1.00
Introduction	0.03-1.00	230 (9.7%) the, of, and, in, study, from, between, these, however	477 (20.2%) literature, et, al, attention, little, whether, known, practices	1035 (44.0%) status, phase, theories, supply, induced, modes, industry, intensity,	609 (25.8%) political, language, cell, academic, dental, deformation
Methodology	0.02-0.99	284 (8.4%) and, the, with, number, data, method, used, was, analysis, following	757 (22.3%) samples, information, performed, calculated, obtained, variable	1642 (48.5%) temperature, respondents, surface, protein, incubated, volume, pressure, gender	697 (20.6%) students, firm, buffer, corpus, steel, election, soil, vertical, purified, crack, texts
Results	0.02-0.94	297 (10.0%) and, the, significantly, table, different, higher, average, may	803 (27.0%) interaction, respectively, relationship, due, example, range, seen, per, evidence	1290 (43.4%) surface, region, fit, stress, density, items, pressure, wave, perceived, concentrations	580 (19.5%) fracture, protein, precipitation, thickness, race, solar, amplitudes
Discussion	0.02-0.99	363 (11.5%) the, and, from, explanation, consistent, possible, associated, in, line, supported	731 (23.2%) findings, et, al, quality, implications, determined, improve, contributes, enhance	1349 (42.9%) performance, social, risk, complexity, density, article, transfer, ratio, detected,	701 (22.2%) strain, English, reading, mechanical, angle, ion, writers, bond, affinity

As can be seen from Table 4.4, DP_{norm} values for high-frequency words in the IMRD sections vary substantially. On average, more widely dispersed words were found to accommodate about 30 percent of high-frequency word types across the sections (M= 33.07%), whereas more narrowly dispersed words accounted for more than three-fifth of these words (M= 66.72%). These results indicate a considerable degree of unevenness in word frequencies across the four sections of research articles.

More differences can be identified by looking at these results in light of disciplinary preferences for the use of widely and narrowly dispersed words across the IMRD sections. For this purpose, the lists of high frequency words from the four sections of research articles across the ten disciplines were analyzed in terms of the extent to which they included items from the two dispersion categories. Tables 4.5-4.8 present the percentages of each category in the IMRD sections across disciplines.

Table 4.5. *Proportion of dispersion categories across disciplines in the introduction section*

	% Widely dispersed	% Narrowly dispersed
Biology	64.29	35.32
Chemistry	64.52	35.18
Dentistry	61.66	38.00
Mechanical Engineering	59.79	40.21
Physics	57.94	39.97
Mean	61.64	37.74
Minimum	57.94	35.18
Maximum	64.52	40.21
Applied Linguistics	65.50	34.50
Business	66.79	33.21
Management	66.96	33.04
Politics	68.56	31.44
Sociology	67.60	32.40
Mean	67.08	32.92
Minimum	65.50	31.44
Maximum	68.56	34.50

Table 4.6. *Proportion of dispersion categories across disciplines in the methods section*

	% Widely dispersed	% Narrowly dispersed
Biology	48.47	51.53
Chemistry	42.04	57.95
Dentistry	45.45	54.55
Mechanical Engineering	51.06	48.94
Physics	54.24	45.76
Mean	48.25	51.75
Minimum	42.04	45.76
Maximum	54.24	57.95
Applied Linguistics	74.48	25.52
Business	74.86	24.92
Management	77.50	22.27
Politics	77.63	22.37
Sociology	76.89	23.11
Mean	76.27	23.64
Minimum	74.48	22.27
Maximum	77.63	25.52

Table 4.7. *Proportion of dispersion categories across disciplines in the results section*

	% Widely dispersed	% Narrowly dispersed
Biology	69.47	30.53
Chemistry	71.37	28.63
Dentistry	75.01	24.99
Mechanical Engineering	66.08	33.92
Physics	62.45	37.55
Mean	68.88	31.12
Minimum	62.45	24.99
Maximum	75.01	37.55
Applied Linguistics	67.24	32.76
Business	73.32	26.68
Management	81.13	18.87
Politics	70.75	29.25
Sociology	69.19	30.81
Mean	72.33	27.67
Minimum	67.24	18.87
Maximum	81.13	32.76

Table 4.8. *Proportion of dispersion categories across disciplines in the discussion section*

	% Widely dispersed	% Narrowly dispersed
Biology	70.15	29.85
Chemistry	68.54	31.46
Dentistry	66.95	33.05
Mechanical Engineering	65.30	34.70
Physics	63.70	36.30
Mean	66.93	33.07
Minimum	63.70	29.85
Maximum	70.15	36.30
Applied Linguistics	67.27	32.73
Business	66.89	33.11
Management	67.69	32.31
Politics	71.95	28.05
Sociology	69.69	30.31
Mean	68.70	31.30
Minimum	66.89	28.05
Maximum	71.95	33.11

As can be seen for the Tables 4.5-4.8, lists of high frequency words from introduction, results and discussion sections in hard and soft disciplines contained similar proportions of individual words from the two categories. However, these proportions were found to substantially vary in the methodology section. As shown in Table 4.6, on average, more than half of high-frequency word types in methodology sections in hard science research articles fall into the category of narrowly dispersed words (M= 51.75%). This is twice the proportion of only 23.64% found for methods sections in the soft fields. In contrast, widely dispersed words were found to be more present in soft field methods, accounting for more than 75% of frequent individual words in this section. These results indicate that explaining research methods in hard knowledge fields demands a vast repertoire of a type of vocabulary which has a somewhat narrow range of use in academic research writing. Table 4.9 provides examples of this vocabulary employed in the methodology sections of hard science research papers in the DCRA.

Table 4.9. *Examples of narrowly dispersed vocabulary in the methods sections of science research papers*

	Narrowly dispersed	
	Moderately unevenly dispersed 0.5–0.749	Extremely unevenly dispersed 0.75–1.00
Biology	temperature, incubated, protein, surface, site, diameter	purified, acetate, buffer, binding, diluted, centrifuged, resuspended,
Chemistry	energy, cells, protein, acid, reactions, concentration	buffer, species, equation, plasmid, ion, spectra, diluted, residue
Dentistry	acid, incubation, antibody, resin, temperature, protein, diameter	specimen, dental, implant, periodontitis, plaque, tissues, saliva
Mechanical Engineering	pressure, volume, velocity, temperature, displacement, laser	thickness, load, steel, equation, simulations, geometry, deformation
Physics	density, energy, waves, chamber, dynamics, intervals, spatial	flux, pulse, simulations, equation, spectra, solar, zonal, zenith, electrons

As Table 4.9 shows, there appears to be a certain degree of overlap in the use of narrowly dispersed words in methodology sections in these disciplines. To obtain a quantifiable measure of such commonality, the degree of overlap in the use of key words was calculated for each of the four sections (i.e., IMRD) across disciplines. To this end, separate listings of keywords for the IMRD sections across disciplines were generated based on the inclusion criteria for keywords mentioned earlier. Following this, the percentage overlaps between the keyword lists of each section across the ten disciplines were computed. This procedure yielded a symmetric matrix with 780 entries (i.e., $(40 \times 39)/2 = 780$), representing the degree of overlap between 40 keywords lists retrieved from the IMRD sections across ten disciplines (see Appendix A). This procedure was then followed by a cluster analysis to identify potential groupings of disciplines based on the degree of their vocabulary overlap across the IMRD sections of the research article. The analysis was performed using Euclidean distance and average linkage method. Figure 4.2 plots the hierarchical clustering of disciplines based on the degree of overlap in key vocabulary use across the IMRD sections.
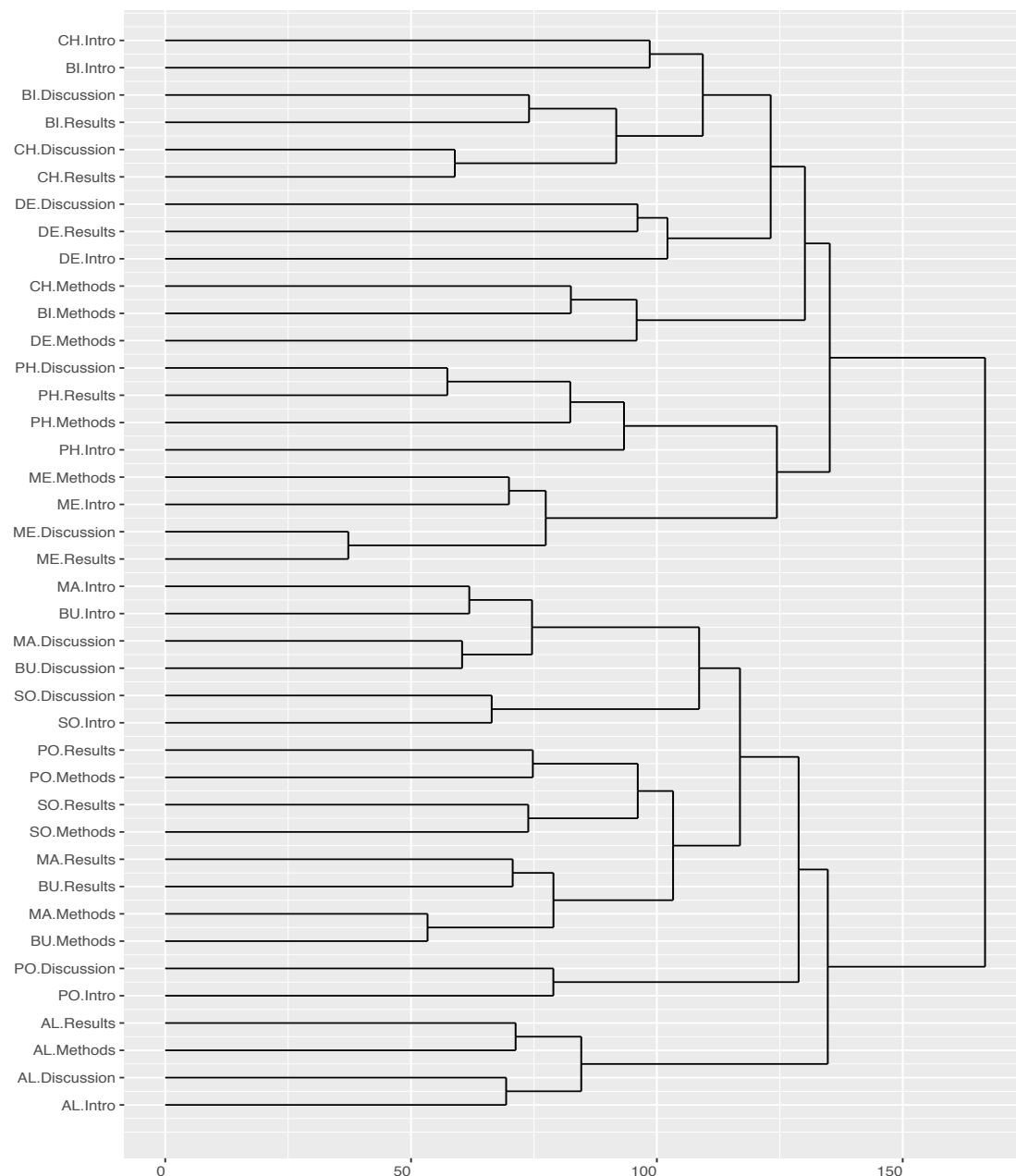
Figure 4.2. Cluster analysis of overlaps between disciplines across the IMRD sections.
BI: Biology CH: Chemistry DE: Dentistry ME: Mechanical Engineering PH: Physics AL: Applied Linguistics
BU: Business MA: Management PO: Politics SO: Sociology

Looking at Figure 4.2, the first point that stands out is that the clustering tree created by the

analysis is divided into two main clusters: one belonging to the IMRD sections in the hard

disciplines (top half) and the other representing the four sections in the soft fields (bottom

half). Reading the top half of the dendrogram, we can see that there are two main groups

comprising the hard sciences cluster: one created by the IMRD sections in mechanical

engineering and physics and the other formed by the sections in biology, chemistry and

dentistry. As can be seen from Figure 4.2, each of these two groups is also composed of certain (sub)clusters. If we look at clusters in the former group, we find that results and discussions in mechanical engineering papers were among the first sections to form a merged pair. Interestingly, introductions and methods in this discipline created their own cluster, indicating a high degree of overlap between these two sections in mechanical engineering. At the next level of grouping for mechanical engineering, the results-discussions pair joins up with the introductions-methods cluster to represent key vocabulary use in this discipline. For physics, results and discussions were also the first to create their own cluster, which then combined with the methods sections in this discipline. This three-point cluster is then joined by a single-point cluster created by introductions in this discipline. The clusters created by the IMRD sections in mechanical engineering and physics then combined to form a stand-alone cluster for these two disciplines, representing a degree of similarity between these two fields in terms of key vocabulary use across the four sections.

Looking at the formation of (sub)clusters in the latter group (biology, chemistry, and dentistry), we can see that, at the first step of clustering in this group, results and discussions in chemistry formed a cluster. This pair is then joined by a cluster created by the same sections in biology, indicating a confluence of these two fields in terms of vocabulary overlap in results and discussions. This four-point cluster is then combined with a merged pair formed by introductions in biology and chemistry. Figure 4.2 shows that, for dentistry, these three sections (introduction, results and discussions) were found to form their own cluster prior to merging with the group assembled by the same sections in biology and chemistry. Further, as shown in the dendrogram, there is a separate cluster for methods in biology, chemistry, and dentistry. This cluster is formed by methods in biology and chemistry creating a pair in the first stage of clustering and then joining up with methods in dentistry. This group of three then combines with the larger cluster formed by introductions, results and

discussions across the three disciplines. This group in turn merges with that forged by the sections in mechanical engineering and physics, representing the broad category of the hard sciences.

Reading the bottom half of the dendrogram, we can see that the grouping of sections in the soft disciplines is rather different from that in the hard sciences. The first point that can be noted is that, unlike the hard sciences, results and discussions were not found to create a pair in any of the soft fields. Instead, it is introductions and discussions that appear to have a high degree of vocabulary overlap in research articles written in the soft disciplines. Interestingly, these two sections were found to make a pair in all the soft fields. A similar trend can also be observed for methods and results. As can be seen from the figure, these two sections were also found to show a high degree of commonality in terms of keywords across all the soft disciplines. Another interesting point is that applied linguistics was once again found to be an outlier among the soft fields. From the Figure 4.2, we can also see that the IMRD sections in applied linguistics separate themselves from those in the other soft disciplines by forging their own group first and then joining the other disciplines in the final round of clustering. This means that the four main sections of research papers in applied linguistics had the lowest average degree of *dissimilarity* with those in the other soft disciplines only when they were grouped together, and not separately. Taken together, these findings lend support to the idea that there are various parameters of variation in vocabulary use across academic disciplines and that the hard vs. soft distinction can only represent one axis of this rather systematic variation. The following section will provide a more in-depth discussion of these differences.

4.4. Discussion

The present study aimed to investigate the commonalities and differences in the language of research writing across disciplines. For this purpose, I examined specificity in the use of single words in the main sections of research articles written in ten different disciplines. In doing so, an inductive approach was adopted through which specific patterns of vocabulary use were systematically identified and analyzed to arrive at a comprehensive picture of disciplinarity in expert research writing. The key strength of this approach is that it allows disciplinary-specific patterns to emerge from the analysis with no *a priori* assumptions guiding their identification. It also allows the researcher to triangulate and verify the emerged patterns using different methods of data analysis. It can be argued that such a methodological triangulation can yield results that can lay greater claim to generalizability, allowing the researcher to anchor findings in more robust theoretical interpretations (see Baker & Egber, 2016; Layder, 1993). Below, I discuss the findings that emerged from the analysis on the basis of the analytical steps taken at different stages of the study.

The first step in the analysis was to assess the degree to which frequently used single words in research writing were evenly dispersed across the disciplines. The results revealed substantial variability (ranging from 0.031 to 0.973) in the dispersion values (DP_{norm}) for high-frequency words used in the corpus, indicating that the language of academic research writing is characterized by highly specialized, field-specific discourses that are far from being homogeneous. These results stand in stark contrast to the theoretical position that views the language of the academy as comprising an array of relatively homogeneous discourses, whose communicative functions are often realized by a set of shared linguistic items (Blue, 1988; Jordan, 1997; also see Hutchinson & Waters, 1987). The analysis showed that about 60 percent of the vocabulary used in the corpus were narrowly dispersed words (i.e., words that were used more frequently in one discipline than the other). These results not only highlight

the importance of this group of vocabulary in the production of scholarly knowledge in research articles, but also show that the majority of the vocabulary used in research writing is not widely shared among academic disciplines and so may well not be of equal value to writing in different branches of academia. It was further found that about two-fifth of individual words in research papers written in the hard disciplines were narrowly dispersed, that is for every five individual words in these disciplines, two had a rather narrow distribution characteristic. Further, the mean percentage of narrowly dispersed words in the sciences was found to be about 1.65 times more than that in the soft disciplines, implying that research writing in the sciences demands the knowledge of a type of vocabulary that has a narrow range of use and applicability. These results corroborate those of Durrant (2014), who found that university students' writing in Science and Technology fields makes greater use of specialized vocabulary, compared to Social sciences and Huminites. The patterns observed in the present study extend Durrant's (2014) findings by indicating that disciplinary differences in the degree of reliance on specialized vocabulary (with narrow range of applicability) is even more pronounced in writing for research publication purposes. This indicates that conventional practices of disciplinary communities can have a profound impact on the lexical choices that members make in this particular genre of writing. Moreover, this finding also provides compelling evidence in support of the argument that regards academic research writing as a highly conventionalized genre in which disciplinary knowledge plays an important role in adhering to the linguistic expectations of the target readership (see Bhatia, 1993; Hyland, 2001; Swales, 1990).

To further explore the observed disciplinary patterns, the next step in the analysis focused on the degree of overlap in the use of key vocabulary between all disciplines in the corpus. A number of key differences and similarities were observed. First, the overlap values between disciplines corresponded to the *hard-soft* classification established in previous

research (Biglan, 1973a; Hyland, 2004; Durrant, 2017). It is important to note that the disciplinary classifications proposed by these studies are based on parameters of variation other than single word vocabulary, with Biglan (1973a) focusing on faculty members' judgements and perceptions of knowledge fields and Hyland (2004) and Durrant (2017) basing their categorizations on the use of MWEs, such as lexical bundles (see Chapter 2, Sections 2.3.1 and 2.3.4). Secondly, the degree of homogeneity in the use of key vocabulary was found to substantially vary even within the hard and soft broad divisions. For instance, mechanical engineering and physics showed a greater degree of overlap with each other than they did with Biology, chemistry, and dentistry. Within the soft category, applied linguistics was found to be an outlier in terms of key vocabulary use, as it showed the lowest average degree of overlap with other soft fields (management, business, sociology, and politics). This provides support for previous research (e.g., Durrant, 2017), which also found a general tendency for humanities to be an outlier in multi-word vocabulary use in university students' writing. The observed patterns of overlap were further verified by the results of a cluster analysis. The cluster analysis provided a systematic mapping of the levels of commonality and variations between disciplines, illustrating that the degree of homogeneity in the use of high frequency words that play a key role in the construction of knowledge in academic disciplines not only varies between hard and soft knowledge fields but also within these two broad categories.

Our further investigation focused on exploring variation in research writing practices in terms of vocabulary use across different sections of research articles. The analysis showed that the dispersion of high frequency words substantially varied across the IMRD sections of research articles. On average, more narrowly dispersed words accounted for more than three-fifth of high-frequency word types across the sections, whereas more widely dispersed words were found to accommodate about 30 percent of these words. This finding indicates that there

is a considerable degree of unevenness in single-word vocabulary use across the main sections of research articles. More differences were identified by looking at these results in light of disciplinary preferences for the use of widely and narrowly dispersed words across the sections. It was found that, compared to writing in Social sciences and Humanities fields, science writing made greater use of narrowly dispersed words across all the four sections. These patterns of results parallel those observed in the first part of the analysis, which revealed science writing's reliance on less-widely used vocabulary, and extend these results by showing that this pattern of use is consistent across all the main sections of research papers. It was also observed that more than half of high-frequency word types in methods sections in hard science research articles were narrowly dispersed across the IMRD sections. This was found to be twice the proportion of only 23.64% found for methods sections in Social sciences and Humanities. This particular observation suggests that explaining research methods in hard knowledge fields demands a vast repertoire of a type of vocabulary which can have a very narrow range of use in academic research writing. These findings also serve to illustrate that disciplinary variation can exist at various levels of specificity (e.g., in the overall text, across the article sections, across different parts of a single section of research articles, see Section 2.3.2; also see Gray, 2015). Indeed, there is no one correct level at which such variation should be identified; rather, it depends on the goals of the researcher (Biber & Conrad, 2009). However, it can be argued that investigating patterns of disciplinary use at different levels of granularity can result in more generalizable conclusions (see, for example, Biber, Egbert, Gray, Oppliger, & Szmercsanyi, 2016).

Further, to explore possible commonalties between sections of research articles in terms of vocabulary use across disciplines, the degree of overlap in the use of high-frequency key words was calculated for each of the four sections across the ten disciplines. It was shown that the degree of overlap between sections is also governed by disciplinary

conventions. More specifically, it was found that introductions and discussions had a high degree of vocabulary overlap in research articles written in Social sciences and Humanities. However, in the sciences, it was results and discussion sections that showed a high degree of commonality in terms of keywords. This finding indicates that there are fundamental differences in how justification of findings and making claims about their significance are handled across disciplinary fields. This provides support for Hyland's (2004, 2008a) assertion that disciplines can characteristically vary in how they convince the reader to assent to a particular interpretation or a knowledge claim in their research papers. The patterns of results revealed in this part of the analysis indicate that the rhetorical purposes of the Discussion section in research articles written in the soft fields are often realized through the lens provided by the arguments already made in the Introduction section. In comparison, the discussion of findings and the claims about their importance in hard science research articles appear to be primarily based on the observations described in the Results section. In other words, while justification of results in Social sciences and Humanities is realized through situating them within a theoretical basis often established in the Introduction sections, discussion of findings and their validity in the sciences is often carried out by reiterating their grounded, experimental basis and anchoring them in empirical observations, rather than theoretical interpretations. It can therefore be argued that creating a convincing discourse for research findings in science writing, in comparison to writing in the soft fields, is more empiricist and less interpretative in nature (more on this in Chapters 5 and 6).

Finally, applied linguistics was, one again, found to be an outlier among the soft fields, with the IMRD sections in this discipline forging their own group first and then joining the other disciplines in the final round of clustering. This suggests that, for writing the main sections of their research papers, writers in the field of applied linguistics would seem to need an inventory of lexis which is not often shared among other soft knowledge fields.

4.5. Conclusion

Academic research writing is one of the most central activities of academic institutions. Scholars in various academic fields share and disseminate the outcome of their scientific endeavor through writing and publishing academic journal articles. It is through this particular form of knowledge dissemination that a rich body of scientific knowledge about a given phenomenon is accumulated. The results presented in this study show that the lexical choices academics make at different stages of composing their research papers is differentially affected by the standards and conventions of scholarly activities in their field. It was found that such conventions have the potential to govern the delineation of authors' linguistic decisions at the most basic levels, such as the lexis.

The present research also attempted to show the strength of inductive methods and methodological triangulation in highlighting and verifying data-driven patterns of disciplinary writing. Future research could adopt such methodological approaches to shed further light on various aspects of academic research writing across disciplines.

4.6. Acknowledgements

Chapter 5

Study 2: A cross-disciplinary investigation of the form-function mapping of multi-word expressions

5.1. Introduction

Recent years have seen a surge in interdisciplinary research on MWEs (also known as formulaic sequences, formulas, prefabricated units, etc., see Wray, 2002). Various types of MWEs such as collocations, lexical bundles, multi-word verbs, idioms, and binomials have been investigated from a number of different perspectives: computational linguistics and natural language processing (Keller & Lapata, 2003; Manning & Schutze, 1999), prosody, fluency, and intonation (Lin, 2013, 2018), first language acquisition (Cruttenden, 1981; Peters 1983), second language acquisition (Schmidt, 1983; Vihman, 1982), native speaker versus LX learner discourse (Durrant and Schmitt, 2009; Siyanova & Schmitt, 2008), on-line processing and representation (for an overview, see Siyanova-Chanturia & Van Lancker Sidtis, 2019), and classroom pedagogy (Meunier, 2012; Wood, 2002). Research in these areas has been instrumental in improving our understanding of the linguistic, discourse, pragmatic, cognitive, and psycholinguistic aspects of MWEs. One area, in particular, has seen considerable interest in recent years -- the use of MWEs (specifically *lexical bundles*) in written published academic discourse. Lexical bundles are referred to as "the most frequently recurring sequences of words in a register" (Biber, 2009, p. 282). They are normally identified using a quantitative, frequency-based approach, employing a certain frequency cut-off (e.g., 10, 20, 40 occurrences per million running words). And so, the identification of sequences relies solely on their distributional properties and not impressionistic evidence, such as the researcher's intuition or perceptual judgment (see Section 2.3.4.2.). Another important criterion for the identification of this particular kind of MWEs is the length of the

bundle in question, for example, 3-, 4-, or 5-word bundles. Save for a few exceptions (Cortes, 2013; Omidian, et. al., 2018), most studies to date have focused on 4-word bundles, as they are often frequent enough to provide a sufficient amount of data and are likely to play more definable functional roles than shorter bundles (Biber, Conrad, & Cortes, 2004; Chen & Baker, 2010; Durrant, 2017; Hyland, 2008).

While frequency can be said to be the ultimate quality of lexical bundles (Cortes, 2015), the two characteristics of these sequences that have been of particular interest to researchers in the context of academic writing are the communicative functions of bundles and their sensitivity to variation in text types (e.g., see Biber, et al, 2004; Chen & Baker, 2010; Cortes, 2004; Durrant, 2017; Hyland, 2008a; 2008b; Omidian, Siyanova, & Durrant, 2020). An emerging trend in this line of inquiry is an emphasis on the connection between lexical bundles and the rhetorical functions they often perform in different parts of texts across different disciplinary fields (Cortes, 2013; Le & Harrington, 2016; Lu, et. al, 2020; Gray, Cotos, Smith, 2020; Omidian et.al, 2018). The overarching aim of this emerging body of research is to address what has recently been labeled by Moreno and Swales (2018, p. 3) as "the function-form gap" in EAP writing and determine whether situational parameters, such as disciplinary conventions, can influence the correspondence between linguistic forms (in this case lexical bundles) and their communicative functions.

So far, this line of research has provided compelling evidence that the conventions and particularities that characterize the epistemological orientations of an academic community can have a profound impact on the form-function connection between lexical bundles and the communicative aims they commonly fulfil. It has also been shown that lexical bundles have the potential to clearly delineate the boundaries between disciplines and show how authors adopt different discursive strategies to promote their research papers within their respective academic/scientific domain (e.g., Hyland, 2008a). The findings from

these studies suggest that a close investigation of lexical bundles and their associated rhetorical functions in high stakes genres of academic writing, such as the research article, could highlight important distinctions between disciplines.

However, studies adopting such a function-oriented approach to researching lexical bundles (and linguistic items in general) are still quite rare (see Lu, et.al, 2020). In addition, the primary emphasis of the extant body of research in this area has been on the micro structures (rhetorical moves, Swales, 1991) of a particular section of research articles (e.g., Introduction, Cortes, 2013 and Lu, et. al, 2020; Discussion, Lee & Harrington, 2016; Abstract Omidian, et al, 2018); and so little attention has been paid to variation in the form-function mappings of linguistic patterns in different sections of journal papers (i.e., IMRD sections) across academic disciplines. To bridge this gap, the present study set out to investigate disciplinary variation in terms of lexical bundles commonly used to fulfill the communicative aims of different sections of research articles. For this purpose, a bottom-up approach was adopted in which lexical bundles were systematically identified and subsequently analyzed in relation to the section and the disciplines in which they were observed. The following section describes the methods used in the present study in more detail.

5.2. Methodology

5.2.1 Identification of lexical bundles

Lexical bundles are the most frequently occurring type of MWEs that function as basic building blocks of discourse (Biber, Conrad & Cortes, 2004). In order for a MWE to be considered a lexical bundle, it should meet certain frequency and dispersion criteria (Biber et al, 1999). Lexical bundles in this study were identified using a computer program developed by the researcher in Perl (version 5.010). The program was designed to look for 4+-word recurring n-grams (i.e., contiguous sequences of n words). The frequency thresholds for the

identification of the expressions were set according to the length of the sequences. Since length inversely correlates with frequency (the longer a bundle is the more likely it is to be of lower frequency), the moderately high frequency cut-off point of 20 occurrences per million words (pmw) was employed for identifying 4-word lexical bundles, ten times pmw for 5-word bundles, seven times pmw for 6- and 7-word bundles, and five times pmw for longer bundles (see Cortes, 2013; Omidian, et al, 2018). In addition, to ensure that the identified bundles were typical of the entire corpus and not restricted to a few texts or certain writing styles, the cut-off point of 5+ texts was employed for range. That is, only those sequences that occurred in five or more texts were included in our list. Finally, for shorter sequences embedded within longer bundles, only those that met the mentioned inclusion criteria were included in the final list. For instance, the six-word n-gram *the results of this analysis are* would only be considered as a lexical bundle if it was observed seven times per million words by itself and not part of the 9-word bundle *the results of this analysis are presented in table*. The final list included 1416 lexical bundle types (i.e., the number of different bundles) and 50,863 tokens (i.e., the total number of bundles).

5.2.2 Lexical bundles in different sections of research articles

All the bundles on the final list were examined in terms of their communicative roles in different sections of RAs. For this purpose (and in line with earlier research), the identified bundles were classified according to their structural correlates using the structural classification scheme proposed by Biber, et al, (2004). For the sake of comparison, the bundles were grouped under three broad categories (as has been done in previous studies, Chen & Baker, 2010, 2016; Cortes, 2013; Pan, Reppen, & Biber, 2016). These groupings included (i) noun phrases (NP-based) (ii) prepositional phrases (PP-based) (iii) bundles incorporating verb phrase fragments (VP-based) (see Table 5.1). Similar to Cortes (2013), a

new category was also created to accommodate bundles comprising both noun phrases and verb phrases (fragments or clauses). Table 5.1 presents the final categories, along with a random selection of bundles from each grouping.

Table 5.1. *Structural classification scheme (adopted from Biber et al, 2004)*

NP-based bundles (NP)
noun phrases incorporating *of*-fragments
(*a comparison of the, the nature of the*)
noun phrases with post-nominal clause fragments
(the extent to which, the degree to which)

PP-based bundles (PP)
prepositional phrase expressions
(*as a function of the, at the same time*)
comparative expressions
(*as well as the, to a greater extent)*

VP-based bundles (VP)
bundles incorporating passive verbs + prepositional phrase fragments
(*are summarized in table, is based on the*)
bundles incorporating passive verbs + dependent clause fragments
(*was used to determine whether*, *has been suggested that the*)
bundles incorporating non-passive verbs + *that*-clause fragments
(*these results suggest that the, the results indicate that the*)
bundles incorporating anticipatory *it* fragments + verb/adjective + (complement clause)
(*it is well known, it is possible to)*

NP- and VP-based bundles (NP-VP)
bundles incorporating both noun and verb phrase fragments
(*the results of this analysis are presented in Table, the aim of this study was to*)

Following this, the functional taxonomy proposed by Biber et al (2004) and Hyland (2008b) was used to categorize the bundles according to their general functions in discourse. In order to better accommodate the identified bundles, certain minor adjustments were made to these classifications. Specifically, since no examples of Hyland's (2008) *topic* bundles were found, this sub-category was excluded from the *research-oriented* category and was replaced with the *intangible framing attributes* sub-category in Biber et al's (2004) classification scheme. This was deemed necessary as certain bundles were found to serve the

function of describing abstract and intangible characteristics of the noun that followed (e.g.,

the validity of the questionnaire). In addition, the *stance feature* sub-category in Hyland's

classification was divided into the subcategories of *centrality, epistemic* and

attitudinal/modality stance bundles to better distinguish between these two functions. The

final categories are presented in Table 5.2.

Table 5.2. *Functional classification scheme (adapted from Biber et al, 2004; Hyland, 2008)*

Research-oriented bundles – provide a structure for writers to explain their research activities and experiences of the real world. Subcategories include:

Location – Indicating time/place
(*at the time of data collection, during the course of the*)

Procedure – Indicating the ways in which experiments and research were conducted
(*the use of the, included in the model*)

Quantification – Indicating quantities or amounts
(*an increase in the number of, a broad range of*)

Description – Indicating the physical characteristics of the following noun
(*the size of the, the crystal structure of the*)

Intangible framing attributes – Indicating abstract characteristics of the following noun
(*the impact of the, the complexity of the*)

Text-oriented bundles – help writers to organize the text and its meaning as a message or argument. Subcategories include:

Transition signals – Establishing additive, contrastive, or equivalence relations between elements
(*in addition to the, on the other hand*)

Resultative signals – Marking causative or inferential relations between elements
 (*as a result of, findings suggest that the*)

Structuring signals – Indicating the organization of different stretches of discourse or referring the reader to specific parts of the text
(*in the present paper, are given in figure*)

Framing signals – Specifying the scope of arguments or situating them within a specific context
(*beyond the scope of this study, in the context of the*)

Participant-oriented bundles – convey the writer's interpretation of a given proposition

Centrality – Indicating the centrality and importance of something
(*one of the key, is one of the most important*)

Epistemic stance – Indicating epistemic judgements and evaluations of a given proposition in terms of its certainty/uncertainty
(*are more likely to, less likely to engage in*)

Attitudinal/modality stance – Indicating the writer's attitude towards the possibility of something
(*it is possible to, can be attributed to the*)

Engagement – Engaging the reader in the process of interpretation
(*it is well known that, it is evident that the*)

Once the bundles were grouped according to their structures and functions, a separate Perl program was written to determine the occurrence rates of each bundle type across the main four sections of RAs (i.e., IMRD) in each discipline. Separate counts were generated for each of the structural and functional categories. These counts were then normalized (per million words) to eliminate the effect of text length. In the following sections, I describe the quantitative and qualitative analyses carried out based on the frequency of occurrences of the identified bundles across the IMRD sections of RAs in different disciplines.

5.3. Analysis and results

5.3.1 Quantitative analysis

5.3.1.1 Distribution of bundles across IMRD sections

As was mentioned above, the identification procedure resulted in a total of 1416 bundle types and 50,863 tokens. Table 5.3 provides information about the number of bundle types and tokens found in each discipline.

Table 5.3. *Normalized frequency of bundle types and tokens used in each discipline.*

Disciplines	Types	Tokens	Type/Token
Biology (BI)	2457.2	10185.4	0.24
Chemistry (CH)	2938.3	13027.6	0.22
Dentistry (DE)	3500.4	12387.2	0.28
Mechanical Engineering (ME)	2664.6	16727.4	0.15
Physics (PH)	3155.7	13537.5	0.23
Hard sciences	**848.1**	**13105.1**	**0.06**
Applied Linguistics (AL)	2453.6	12826.2	0.19
Business (BU)	2283.6	13813.4	0.16
Management (MA)	2366.0	13460.6	0.17
Politics (PO)	2557.4	13695.2	0.18
Sociology (SO)	2257.1	12910.5	0.17
Soft sciences	**583.4**	**13160.8**	**0.04**

Following their identification, all the bundles were grouped under the sections in which they occurred. Table 5.4–5.5 illustrate the distribution of the identified bundle types and tokens in different sections of RAs across the ten disciplines.

Table 5.4. *Normalized frequency of bundle types in different sections of RAs across disciplines.*

Disciplines	Introduction	Methods	Results	Discussion
Biology (BI)	5749.8	3295.0	4430.9	5666.6
Chemistry (CH)	7846.6	3461.0	5290.8	6494.8
Dentistry (DE)	7302.7	3341.0	5708.4	6963.9
Mechanical Engineering (ME)	7000.7	5273.6	5262.5	6641.7
Physics (PH)	7056.7	5979.6	5414.1	7086.4
Hard sciences	**3524.2**	**1932.1**	**2188.4**	**2642.4**
Applied Linguistics (AL)	6340.0	5015.5	5248.2	6086.4
Business (BU)	7545.3	5201.0	5775.0	6201.2
Management (MA)	7223.6	5447.9	6794.2	5765.1
Politics (PO)	7448.1	6118.0	6026.8	7331.7
Sociology (SO)	7002.2	4884.1	4830.9	7219.9
Soft sciences	**3312.5**	**2240.9**	**2144.5**	**2350.0**

Table 5.5. *Normalized frequency of bundle tokens in different sections of RAs across disciplines.*

Disciplines	Introduction	Methods	Results	Discussion
Biology (BI)	10219.8	7507.8	11278.7	12372.9
Chemistry (CH)	13506.9	7709.7	15220.8	15400.8
Dentistry (DE)	13004.5	6775.7	13915.9	16121.2
Mechanical Engineering (ME)	13343.7	13028.1	18765.6	18769.1
Physics (PH)	12061.4	10533.6	15177.1	14204.5
Hard sciences	**12325.6**	**8780.0**	**14946.3**	**15392.7**
Applied Linguistics (AL)	11537.8	10623.9	12988.1	15074.6
Business (BU)	13644.4	11983.6	14594.6	15611.5
Management (MA)	12850.2	12307.1	15597.4	14500.0
Politics (PO)	12135.9	12795.8	15506.8	14592.8
Sociology (SO)	11888.5	10999.6	13976.6	15800.2
Soft sciences	**12422.1**	**11730.2**	**14361.1**	**15105.8**

Log-linear modeling was performed to assess whether the distribution of bundles is dependent on the disciplines and IMRD sections in which they were found. As a type of generalized linear regression, Log-linear models use logarithmic link function to examine dependency relations and association patterns between variables (see Gries, 2013, p. 324; von Eye, Mun, & Mair, 2011). Two log-linear (Poisson) models were constructed in R version 3.5.1 (2018-07-02) using the GLM function and car package (version 2.1-5). Each model was built with a different dependent variable, namely, type distribution (Model 1) and token distribution (Model 2) of the identified bundles. Two categorical variables were used as predictors in each model: (i) DISCIPLINE with two levels (Hard and Soft) (ii) SECTION with four levels (Introduction, Methods, Results and Discussion). Since the main purpose of the analysis was to determine whether there is an interaction between DISCIPLINE and SECTION for the use of bundle types and tokens, an interaction term (DISCIPLINE × SECTION) was also included in each model. The results of the models are shown in Table 5.6. As can be seen, the models showed no significant main or interaction effects of DISCIPLINE and SECTION on the type and token frequencies of the bundles.

Table 5.6. *Type-III analysis of variance for the models with DISCIPLINE and SECTION as predictors*

Model	Outcome	Predictor	*LR Chisq*	*df*	*Pr(>Chisq)*
Model 1	Type frequency	DISCIPLINE	8.94	1	.446
		SECTION	95.24	3	.171
		DISCIPLINE × SECTION	12.49	3	.092
Model 2	Token frequency	DISCIPLINE	90.73	1	.136
		SECTION	253.87	3	.087
		DISCIPLINE × SECTION	395.54	3	.064

5.3.1.2 Distribution of structural categories across sections

As was described earlier, the identified bundles were classified according to their structural correlates using the structural classification scheme described in Section 5.2.2 (see Table 5.1). Table 5.7 presents type and token frequencies of each structural category across the IMRD sections.

Table 5.7. *Distribution of structural categories across the IMRD sections*

			NP	PP	VP	NP-VP
Hard	Introduction	Type	1049.8	1017.5	966.6	481.0
		Token	3783.2	3584.3	3445.6	1475.3
	Methods	Type	540.3	604.1	516.3	540.3
		Token	2323.4	2861.0	2166.4	2323.4
	Results	Type	584.6	642.0	650.8	306.6
		Token	3816.5	4535.7	3812.1	2759.8
	Discussion	Type	689.0	753.2	827.3	368.0
		Token	3978.5	4815.6	4949.0	1644.7
Soft	Introduction	Type	895.9	967.4	986.2	440.4
		Token	3764.3	4167.1	3147.0	1027.7
	Methods	Type	601.4	641.5	672.7	309.6
		Token	3648.7	3962.8	2755.5	1185.1
	Results	Type	566.8	598.4	648.6	314.1
		Token	4296.5	4391.2	3779.8	1728.2
	Discussion	Type	588.0	661.8	771.4	314.0
		Token	4253.3	4588.5	4761.3	1325.7

To explore possible variation in type and token distributions of each structural category across the IMRD sections in hard and soft disciplines, two Log-linear models were constructed with three categorical variables as predictors: (i) DISCIPLINE with two levels (Hard and Soft) (ii) SECTION with four levels (Introduction, Methods, Results and Discussion), and (iii) STRUCTURE with four levels (NP, PP, VP, NP-VP). The interactions between these variables were also added to the models. The 'drop1' function in R was then used to identify and remove from the models those interaction terms that were not significant (see Gries, 2013a, p. 266). Table 5.8 presents the results of these models.

Table 5.8. *Type-III analysis of variance for the models with* DISCIPLINE, SECTION, *and* STRUCTURE *as predictors*

Model	Outcome	Predictor	LR Chisq	df	Pr(>Chisq)
Model 1	Type frequency	DISCIPLINE	22.3	1	.046
		SECTION	290.7	3	.085
		STRUCTURE	318.9	3	.03
		DISCIPLINE × STRUCTURE	11.6	3	.043
Model 2	Token frequency	DISCIPLINE	9.19	1	.037
		SECTION	561.5	3	.092
		STRUCTURE	2128.4	3	.033
		DISCIPLINE × STRUCTURE	45.7	3	.026
		DISCIPLINE × SECTION × STRUCTURE	112.7	9	.012

The results of Model 1 (type frequency) showed no three-way interaction between

DISCIPLINE, SECTION, and STRUCTURE (DISCIPLINE × SECTION × STRUCTURE). However, as can

110

be seen from Table 5.6, the interaction between DISCIPLINE and STRUCTURE (DISCIPLINE ×

STRUCTURE) was found to be significant. Since making inferences about main effects in the

presence of an interaction effect could be misleading (cf. West, et al, 2014), I base my

interpretation on the results of the observed interaction effect (i.e., DISCIPLINE × STRUCTURE).

The patterns of use plotted in Figure 5.3 shows that, in general, RAs written in the hard fields

contained more NP-based bundle types than those in the soft disciplines. The results of post-

hoc comparisons showed that the observed difference was significant ($z=3.59$, $p=0.009$),

suggesting that science research writing in the hard fields tends to heavily rely on different

types of lexical bundles with noun phrase structures.
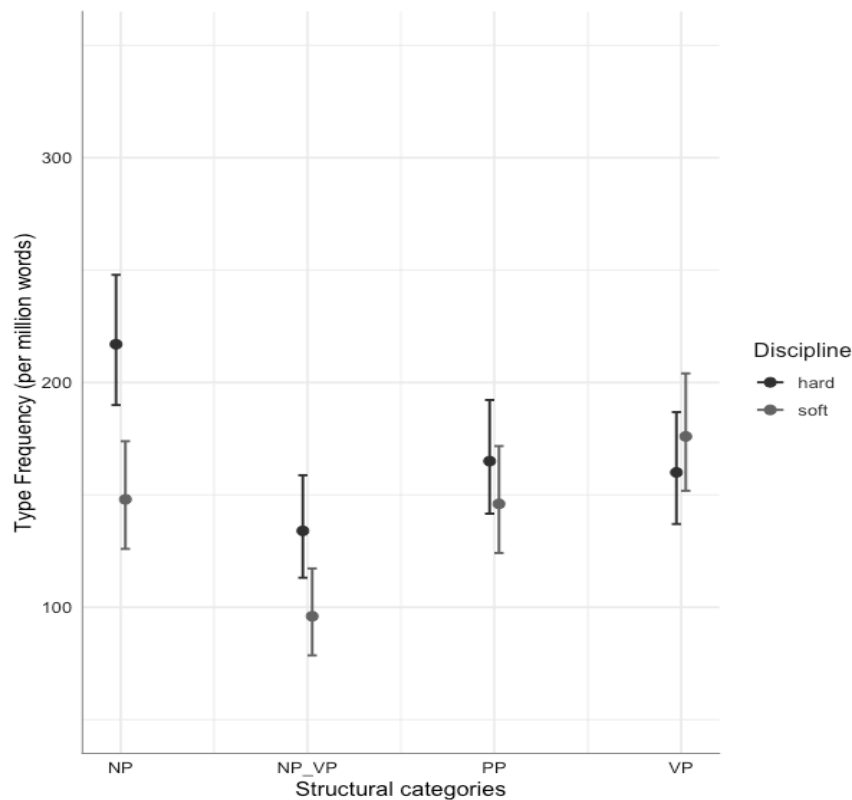
Figure 5.3. Type distribution of structural categories in the IMRD sections across hard and soft disciplines.

The results of the model with the token distribution of structural categories as the

outcome variable showed a significant three-way interaction between the predictors

(DISCIPLINE × SECTION × STRUCTURE). As was noted above, since making inferences about low order terms (DISCIPLINE × STRUCTURE) in the presence of a high order term (DISCIPLINE × SECTION × STRUCTURE) could be misleading, the interpretation of the obtained results should be based on the highest order term in the model (i.e., DISCIPLINE × SECTION × STRUCTURE). The results of this three-way interaction effect are illustrated in Figure 5.4.
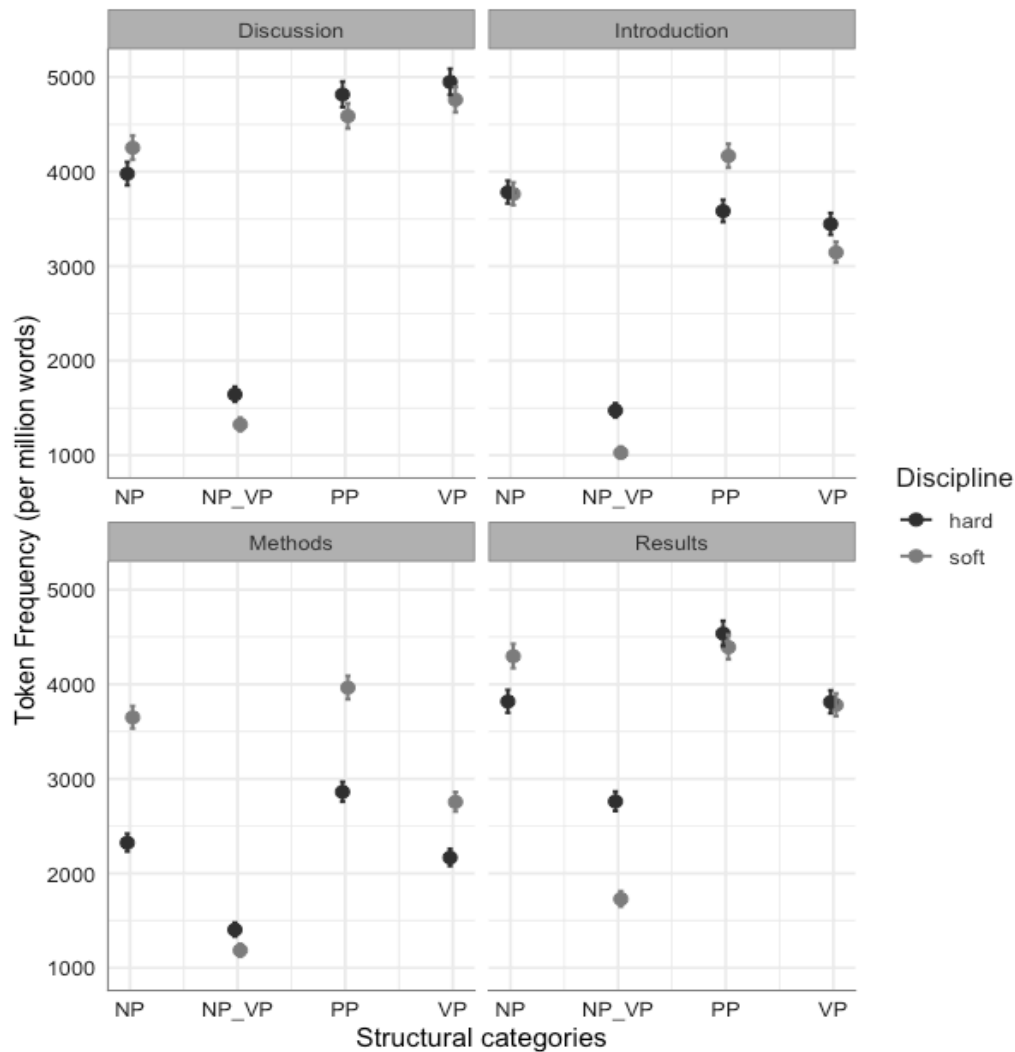
Figure 5.3. Token distribution of structural categories in the IMRD sections across hard and soft disciplines.

A number of points stand out from Figure 5.3. First, the token frequencies of NP-based bundles appear to substantially vary between hard and soft fields in certain sections. More specifically, post-hoc analyses showed that there were more tokens of this type of bundles in

Methods and Results sections in the soft fields compared to the hard disciplines (z=-17.01, p<.001 and z=-5.31, p=.011, respectively).

Second, all the four main sections in RAs written in the hard sciences contained substantially more tokens of NP-VP bundles in comparison to the soft fields. Post-hoc testing showed that all these pair wise differences were statistically significant: Introduction (z=8.88, p<.01), Methods (z=4.28, p=.019), Results (z=15.26, p=<.001), and Discussion (z=5.84, p=.012). As was mentioned earlier (Section 2.2), NP-VP bundles are bundles comprising both noun phrases and verb phrases (fragments or clauses). Structurally, these bundles are often complete units which carry specific and readily discernable communicative functions (e.g., *there were no statistically significant differences*). As illustrated in Figure 5.3, the frequency of occurrence of these bundles is considerably higher in Results compared to the other three sections. This suggests that the frequent use of NP-VP bundles has become conventionalized in the Results section of RAs, particularly those written in the hard fields. Third, certain fundamental disciplinary variation was also found in the use of PP-bundles across sections, with Introductions and Methods in the soft knowledge fields containing significantly more instances of these bundles, compared to those in the hard sciences (z=-6.61, p<.01 and z=-13.28, p<.001, respectively). PP-based bundles are often used to indicate logical relations between propositions (e.g., *as well as the*, *in such a way*). These results indicate that the use of such bundles is considerably frequent in Introduction and Method sections of RAs written in the soft fields.

Finally, with regards to the token distribution of VP-based bundles, the results of post-hoc testing showed that the frequency of occurrence of these verb-based bundles was significantly higher in Methods in the soft disciplines compared to those in the hard sciences (z=8.37, p<.01). This suggests that writers in the soft field tend to rely on verb-based phrases when describing the methodology used in their research.

5.3.1.3. Distribution of functional categories across sections

The identified bundles were also grouped together based on their discourse functions using the functional classification scheme described in Section 5.2.2 (see Table 5.2). Table 5.9 presents the distribution of functional categories across the IMRD sections in hard and soft fields.

Table 5.9. *Distribution of functional categories across the IMRD sections*

			Research-oriented	Text-oriented	Participant-oriented
Hard	Introduction	Type	1503.1	1378.2	638.2
		Token	4805.4	5374.3	2113.6
	Methods	Type	1083.2	590.8	255.5
		Token	5517.1	2584.2	665.4
	Results	Type	1001.6	783.2	403.7
		Token	6677.9	6097.7	2170.8
	Discussion	Type	1113.8	975.5	553.2
		Token	5751.6	6294.9	3346.3
Soft	Introduction	Type	1291.2	1306.2	700.2
		Token	4208.5	5714.2	2205.9
	Methods	Type	1035.8	779.6	416.6
		Token	6551.2	3597.5	1459.0
	Results	Type	936.6	756.3	440.4
		Token	5961.5	5392.9	2884.1
	Discussion	Type	904.2	893.7	543.8
		Token	4255.4	6887.9	3817.0

Log linear models were built to examine possible variation in type and token distributions of each functional category across the four sections in hard and soft fields. The models were constructed with three predictors: (i) DISCIPLINE with two levels (Hard and Soft) (ii) SECTION with four levels (Introduction, Methods, Results and Discussion), and (iii) FUNCTION with three levels (Research-oriented, Text-oriented, and Participant-oriented). The interactions between these predictors were also included in the models. The 'drop1' function in R was once again used to discard those interaction terms that were not significant. The results of the final models are presented in Table 5.10.

Table 5.10. *Type-III analysis of variance for the models with DISCIPLINE, SECTION, and FUNCTION as predictors*

Model	Outcome	Predictor	*LR Chisq*	*df*	*Pr(>Chisq)*
Model 1	Type frequency	DISCIPLINE	0.074	1	.782
		SECTION	195.22	3	.018
		FUNCTION	207.43	2	.013
		DISCIPLINE × FUNCTION	7.41	3	.024
		DISCIPLINE × SECTION × FUNCTION	16.98	6	.009
Model 2	Token frequency	DISCIPLINE	3.083	1	.071
		SECTION	195.93	3	.017
		FUNCTION	102.14	2	.052
		DISCIPLINE × FUNCTION	27.49	2	.039
		DISCIPLINE × SECTION × FUNCTION	19.59	6	.002

The results of both models showed a significant three-way interaction between the predictors (DISCIPLINE × SECTION × STRUCTURE) (Table 5.10). I focus my interpretations on the results of these particular interaction effects, as they are the highest order terms in the models. Figures 5.4 plots the results of the three-way interaction effect for Model 1 (type distribution).
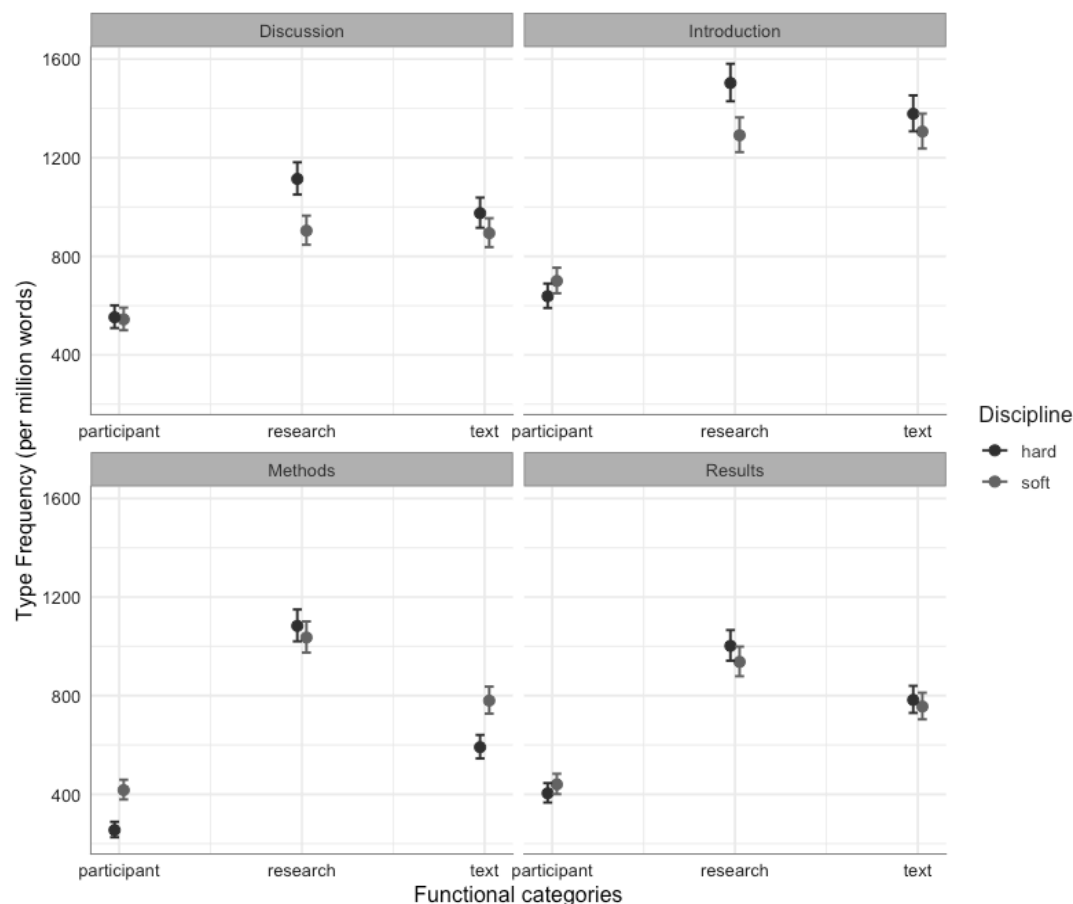
Figure 5.4. Type distribution of functional categories in the IMRD sections across hard and soft disciplines.

The first key point that stands out from the Figure 5.4 is the difference between hard and soft fields in their reliance on using different types of bundles from the research-oriented category when writing Introductions and Discussions. Post hoc testing indicated that these two sections in the hard sciences contained significantly more research-oriented bundle types, compared to those in the soft fields (Introduction, $z=4.66$, $p=<.01$; Discussion, $z=4.007$, $p=.017$). Secondly, authors in the soft disciplines were found to make a greater use of bundle types from text- and participant-oriented bundles when explaining the methodology of their research (text-oriented bundles, $z=5.088$ $p=<.01$; participant-oriented bundles, $z=-6.187$ $p<.01$).

Post-hoc analyses of the three-way interaction effect for Model 2 (token distribution) also revealed a somewhat similar trend. More specifically, the frequency of occurrence of

research-oriented bundles in Discussions written by writers in the hard fields was found to be significantly higher compared to the soft disciplines (z=-6.187 *p*<.01). In addition, text- and participant-oriented bundles were also found to be significantly more frequent in the Methods sections of journal papers in the soft fields than the hard sciences (text-oriented bundles, z=-4.084 *p*=.012; participant-oriented bundles, z=-5.279 *p*<.01). These patterns of use are illustrated in Figure 5.5.
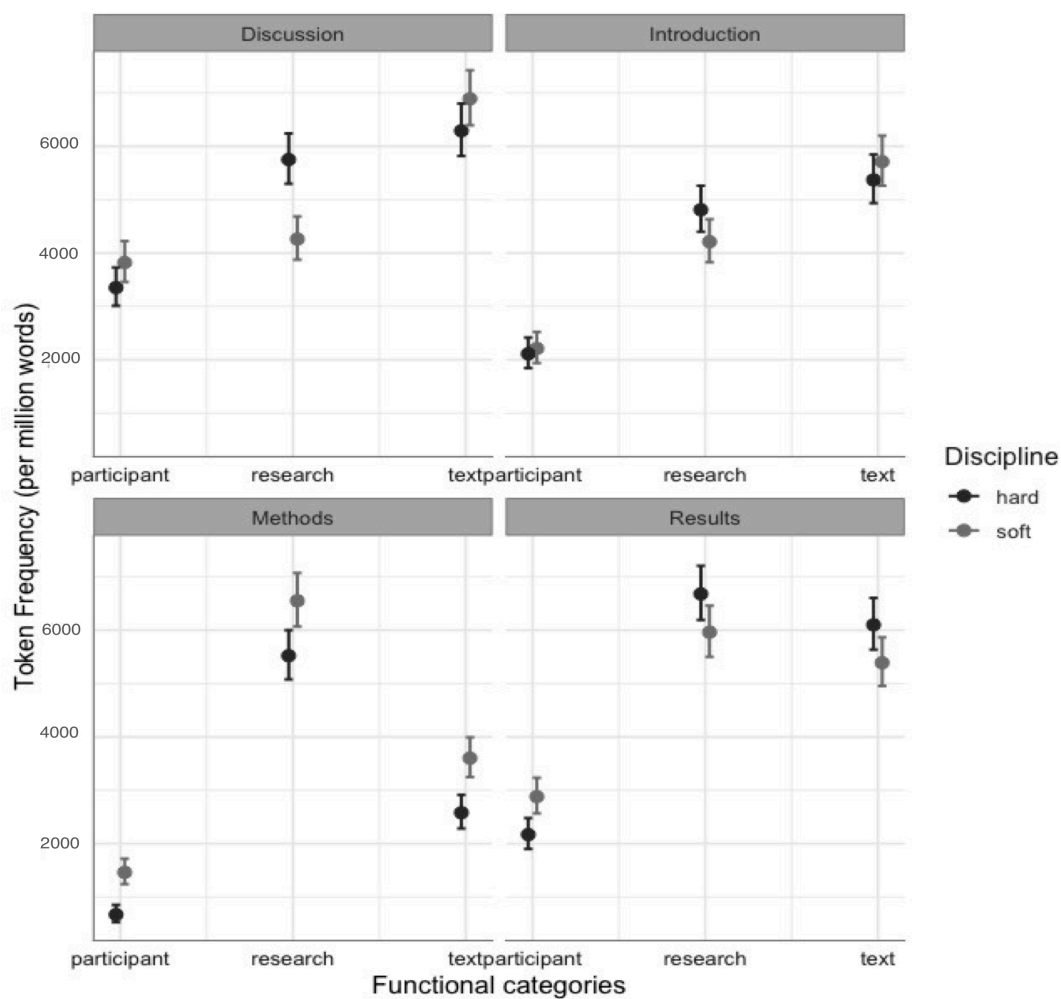
Figure 5.5. Token distribution of functional categories in the IMRD sections across hard and soft disciplines.

Taken together, it would seem that writers in the hard fields tend to use a large number of bundles to describe research activities, equipment and materials when introducing their study and discussing its findings. In contrast, authors in the soft fields appear to have a tendency to use bundles that help organize a form of argument that justifies their research methods. The

117

following section will provide a more in-depth examination of such differences based on a qualitative analysis of the identified bundles in their context of use.

5.3.2. Qualitative analysis

In order to achieve a more fine-grained explanation of the observed quantitative differences, a qualitative analysis of the identified bundles in their surrounding context was deemed necessary. This part of the analysis focused on the usage and communicative functions of the bundles that were found to be distinctive of either hard or soft disciplines in each section of the RA. Distinctive bundles were operationalized as word combinations which were significantly overrepresented in either hard or soft corpus. To identify such distinctive bundles, log-likelihood G^2 test statistic was calculated for each bundle across the IMRD sections. G^2 is one of the most widely used log-likelihood tests for comparing the frequency of occurrence of linguistic items between corpora (see Rayson & Garside, 2000 and Gries & Durrant, 2019). In the present study, G^2 was calculated using the following equation (adopted from Rayson, 2008):

$$G^2 = 2 \times ((O_{corpus1} \times \ln(O_{corpus1}/E_{corpus1})) + (O_{corpus2} \times \ln(O_{corpus2}/E_{corpus2})))$$

where O and E are the observed and expected frequencies of linguistic items in each corpus. The calculation of expected frequencies was performed using the following formulas (ibid):

$$E_{corpus1} = N_{corpus1} \times (O_{corpus1} + O_{corpus2})/(N_{corpus1} + N_{corpus2})$$

$$E_{corpus2} = N_{corpus2} \times (O_{corpus1} + O_{corpus2})/(N_{corpus1} + N_{corpus2})$$

where N is the total number of words in each corpus.

In the following sections, the results obtained from G^2 tests are used as the basis for the qualitative analysis of the distinctive bundles for hard and soft disciplines in various sections of the RA.

5.3.2.1 Introduction section

A comparison of the distinctive bundles in this section revealed key differences between hard and soft science fields (see Appendix B for a complete list of distinctive bundles across sections). First, many of the research-oriented bundles in this section were found to be used by writers in the hard fields to specify the physical attributes of the following noun, describing tangible aspects of its quality, form and size (e.g., *the structure of the, the surface of the*). These nouns were found to be research objects/contexts which were discussed in the article. Many of these bundles were noun phrases incorporating *of*-genitive fragments (e.g., *the mechanical properties of the)*. Excerpts 1 and 2 below show how these bundles are used in context:

(1) ***The structure of the*** peptide bound to their receptors is not clear. However, extensive structure–activity relationship studies have provided the evidence that the C-terminal segment is crucial for the binding of NPY to its receptors [8, 9] (Chemistry)

(2) In the transverse impact loading, the beams are subject to both bending moment and shear force from the punch. The dimensions and shape of the punch will remarkably affect ***the response of the*** beam. The existing theoretical models can in general deal with the beams under pure bending. (Mechanical Engineering)

In addition, hard science writers had a preference for using research-oriented bundles that provide a structure for describing research procedures (e.g., *for the formation of, have been used to, as a function of the*). These bundles were used to provide a detailed explanation of the research processes carried out in previous studies, with the aim of indicating a gap in prior research and developing a cogent argument that emphasizes the centrality of the study being introduced (excerpts 3). In many cases, these bundles were also employed to explain

the operational nuances of the research activities involved in conducting the reported study (excerpt 4).

> 3) Although genomic techniques, such as the microarray analysis of mRNA, ***have been used in*** models of periodontal disease, proteomics has only been applied to the host response in periodontitis in two previous studies… . In this study, we used…
> (Dentistry)
>
> 4) This problem is studied in the framework of tides, which are investigated using daytime data …. Having established the tidal variations, the tide is removed from the WINDII data, and the effects on the perturbations are shown to be small. Finally, the longitudinal variations are studied ***as a function of*** day number, showing that the fall perturbation has a preferred longitude range, which broadens with increasing latitude.
> (Physics)

Certain research-oriented bundles were also found to be distinctive of Humanities and Social sciences writing. These bundles, however, were markedly different from those used in science writing in terms of the discourse functions they commonly perform. More specifically, most of the research-focused bundles found in Introductions of Humanities and Social sciences articles were employed to specify abstract and intangible characteristics of the following noun (e.g., *of the quality of, the nature of the*). It is worth mentioning that about four-fifth of these 'intangible framing' devices were completely absent in hard science Introductions. The following excerpts illustrate the rhetorical work achieved by the use of these bundles in the soft disciplines:

> 6) As business ethics is a growing area of research and publication, an accurate assessment ***of the quality of*** journals of the discipline is helpful for ethics researchers,

their deans and department heads, practitioners, and accrediting agencies (Choi, Kim, & Kim, 2009). Although the previous research… (Business)

7) Given these inconsistent findings regarding the nomological network of the wording factor, it is almost impossible to ascertain its status. For at least three decades now, researchers have been examining **the nature of the** item wording effect; we have yet to understand… (Sociology)

I also examined the use of bundles that were found to be exclusive of the Introduction section, that is they appeared only in this section and were absent in other sections (see Appendix B for a complete list of exclusive bundles in each section). Interestingly, the use of the bundle *the following research questions* was found to be completely absent in all the Introductions written in the hard sciences. A careful examination of this section in hard science RAs revealed that the presentation of research questions is often superseded by a brief description of research equipment, environments, materials, and methodologies. The adoption of this rhetorical strategy appears to allow writers in the hard fields to introduce their study through the medium of its research methodology.

5.3.2.2 Methodology section

The axes of variation between hard and soft science writing in this section were found to be driven by the differences in the use of text- and participant-oriented bundles (also see Section 5.3.1.3). In the case of the former (text-organizers), writers in the soft fields used more than twice as many distinctive bundles from this category as science writers. Humanities and Social sciences writing in this section was particularly marked by the use of distinctive text-organizers functioning as either transitions between different textual elements (e.g., *as well as the, in addition to the*) or framing devices that situate arguments within a specific context (e.g., *with regards to the, in the literature on*). These bundles, which are typically realized by

syntactic structures such as prepositional phrase fragments, made up about four-fifth of distinctive text-organizers found in the Methods sections of Humanities and Social sciences RAs:

> 8) ***In addition to the*** disciplinary groupings, an equal number of award-winning and non-winning finalists' presentations in each of the four selected disciplines were included to … (Applied Linguistics)
>
> 9) A variant of this method has been widely used ***in the literature on*** firm changes (Sociology)

In comparison, the limited number of distinctive text-organizers employed by hard science writers in this section were mostly used as structuring signals to refer the reader to specific parts of the text, such as tables and graphs (e.g., *are given in Table, are shown in Table*). In the case of participant-oriented bundles in this section, all the expressions from this functional category were found to be distinctive of Humanities and Social sciences writing. Many of these bundles were attitudinal/modality expressions employed to indicate and frame the possibility of an action or the forthcoming proposition (e.g., *makes it possible to, more likely to be*). All of these were either completely absent or of very low frequency in the Methods sections of hard science RAs.

> 10) The qualitative method ***makes it possible to*** show how historical, cultural, or political connections in the use of language and communication penetrate a text's content, meanings, structures, or strategies. (Politics)
>
> 11) Firms in a group structure may therefore ***be more likely to*** conduct international innovation. We include a dummy indicating whether the firm is part of a company group. (Business)

As can be inferred from the examples, it appears that writers in soft knowledge fields tend to provide tentative support for their choice of a particular approach/methodology through plausible reasoning. Fulfilling such a communicative function through the use of bundles expressing attitudinal/modality meaning can be taken as a reflection of the writer's need to anticipate possible objections with regards to the methodology employed in their study. It is important to note that such a rhetorical presentation of research methods was not observed in the Methodology sections of RAs written in the hard disciplines.

5.3.2.3 Results section

Authors in both hard and soft fields made relatively widespread use of bundles from all the three functional categories to present the results of their research. A comparison of the distinctive bundles, however, revealed key differences in the use of participant-oriented bundles in this section. On the one hand, Humanities and Social sciences were, once again, found to make extensive use of attitudinal/modality stance markers, such as *be more likely to*, when describing their findings. As illustrated in the following excerpts, it seems that writers in soft knowledge fields tend to avoid overgeneralizations of their results and tone down their views when discussing their research outcomes. This is perhaps due to the divergent nature of the readership in soft disciplines and "fewer unequivocal bases for accepting claims" that writers in these fields tend to ensure that their interpretation of data and results is potentially falsifiable (Hyland, 2008, p. 14).

> 12) While critics ***may be more likely to*** privilege narrowly focused reviews, albums described as diverse by Pitchfork achieve greater market success by this measure. (Sociology)

13) but not reaching the capital as estimates of outcomes of future battles converge. In contexts where there are three or more evenly matched actors, fighting appears ***to be more likely to*** reach the capital. (Politics)

Hard science writing in this section, on the other hand, was found to be marked by the use of participant-oriented bundles functioning as engagement devices, providing a structure for the writer to engage the reader in recognizing the ensuing comment as a reasonable, accepted or accurate view on the issue being discussed:

14) ***It is known that*** some residues (49,50) (K, E, and R, in our study) are present in higher proportions in IDPs than in non-IDPs. (Chemistry)

15) ***It is evident that*** snow cover had a positive correlation with the albedo, while vegetation had a negative correlation with the albedo. In the midlatitudes in Northern Hemisphere, strong negative correlations occurred between SINDVI and albedo. (Physics)

16) ***It can be observed that*** the simulations are proficient at predicting the outwardly-visible residual deformation of the sandwich panel in each case. This is further substantiated by the measured and predicted maximum residual impact dent depth values provided in Table 7. (Mechanical Engineering)

By using these particular bundles, the author attempts to construct a form of solidarity with the reader through appealing to shared knowledge (Hyland, 2008). However, as was stated above, these bundles were found to be distinctive of hard science writing in this section. The relative absence of such bundles in the Results sections of RAs in the soft disciplines can be reflective of the discursive conventions that warrant writers in these fields to carefully marshal their arguments and weigh up the amount of confidence they place in their evaluations and interpretations of findings.

5.3.2.4 Discussion section

Finally, the use of research-oriented bundles in Discussions was found to be distinctive of hard science writing in this section, with writers in these fields employing three times as many distinctive bundles from this category as their counterparts in the soft disciplines. Discussion sections of hard science RAs contained various types of research-focused bundles. These bundles were mostly employed to describe procedures and processes (*in the presence of*), the shape of the material used (*the thickness of the*), and the quantitative aspects of the observations being discussed (*the amplitude of the*). Extracts 17-19 illustrate the usage of these bundles in hard science discussions:

> 17) the grain refinement achieved ***in the presence of*** Al2Gd phase has a major effect in enhancement of the tensile properties of the Mg-4. 8Gd-1. 2Al-1Zn alloy compared with the GZ61 alloy (Mechanical Engineering)
>
> 18) ***The thickness of the*** polymeric film is different in peaks and bottoms of the relief. This causes the difference in strength of the electric field in these film areas. (Physics)
>
> 19) As discussed above, however, both NMR and spectroscopic data do not suggest any significant coupled folding that may account for ***the magnitude of the*** observed ΔC, as secondary structures in the complex are observed in the unbound state as well. (Biology)

Conversely, Humanities and Social sciences writing in this section was marked by the extensive use of text-oriented bundles. More specifically, Discussion sections in these fields contained more than twice as many distinctive bundles from this functional grouping as those in the hard sciences. The overwhelming majority of these expressions in soft science

discussions worked to contextualize arguments, specify conditions under which findings should be evaluated, and provide a unified framework for the interpretation of claims:

> 20) Examining L2 use *from the perspective of* multicompetence highlights the experiences of the L2 users and the ways in which these experiences afford them access to new and varied resources. In this study we were able to focus on how these experiences allowed… (Applied Linguistics)
>
> 21) We were able to indicate this process by showing how it targeted women's sense of entitlement in indirect, implicit ways. It is a violent process *in the sense that* it involves an imposition. (Sociology)

The use of text-organizers in hard science discussions, however, was mostly restricted to bundles pointing to results presented in figures (i.e., *text-structuring signals*) and expressions indicating the inductive reasoning associated with the interpretation of the findings (i.e., *text-resultative signals*):

> 22) This explanation can be confirmed by the strain distribution… Moreover, the area of high strain near the inner hinge for TRTs is…. . For a more intuitive understanding of the differences between these two structures, the subdivision graphs of specific locations *are presented in Fig*. 28. (Mechanical Engineering)
>
> 23) Thus, in principle, (EK) and (E)(K) could lead to very different ensembles. Therefore, the lack of difference in the geometric measures *is due to the* small differences between the EK and KE, and the EE and KK conditionals, as is evident from examination of Table S1 of the Supporting Information (Chemistry)

So, while the discussion of findings in the hard sciences is mostly based on the inductive analysis of observations and results, the explanation of research outcomes in the soft fields is

typically grounded in contextualized interpretations that work to specify the conditions under which research conclusions can be justified.

5.4. Discussion

The present study set out to investigate variation in the form-function mappings of MWEs commonly used in different sections of journal papers (i.e., IMRD sections) across academic disciplines. The main goal of the study was to determine the degree to which writers in hard and soft science disciplines differ in terms of the lexical bundles they commonly employ to realize the communicative purposes of main sections of research articles (i.e., IMRD sections). To explore such differences, a mixed-methods approach was adopted through which the identified patterns of variation were examined both quantitatively and qualitatively. The complementary strength of this hybrid methodology enabled us to uncover a series of differences in focus and practice between hard and soft science fields. Below, I first discuss the patterns of variation in light of the quantitative analysis and then specify some of the major differences uncovered in the qualitative inquiry.

The first step in the quantitative analysis focused on disciplinary variation in the type and token frequencies of lexical bundles (irrespective of their structural or functional classification) across the IMRD sections. The results of Log-linear models showed no significant differences between hard and soft science fields in the frequencies of the types and tokens of bundles used in the four sections. This finding suggests that the overall distributional patterns of lexical bundles are not the strongest predictor of variation in disciplinary writing. In the second phase of the analysis, all the identified bundles were classified according to their structural correlates and the frequency distribution of each structural group was examined in the four sections across disciplines. The results showed that NP-based bundle types (but not tokens) were more common in research articles written in the

hard fields than those in the soft disciplines, suggesting that science research writing tends to employ different types of lexical bundles with noun phrase structures. However, it was also found that writers in the soft disciplines made greater use of NP-based bundle tokens in Methods and Results sections, in comparison to their peers in the hard fields. Given that the type frequency of NP-based bundles was significantly lower in the soft fields than the hard sciences, it can be argued that writers in the soft disciplines tend to make extensive use of a limited range of these bundles when reporting the methodology and results of their study (e.g., *the extent to which the*). Previous research has shown that the use of noun phrase structures are very common in academic prose, indicating that academic writing relies heavily on such structural features, as opposed to other registers (e.g., Biber, et. al., 1999; Cortes, 2004). The patterns of results revealed in the present study extend the findings from these studies by indicating that there are systematic patterns of disciplinary variation in the use of noun phrases across sections of research articles and that the frequent use of such structures in academic (research) writing should be considered in relative rather than absolute terms.

In addition, it was found that writers in the sciences used more tokens of NP-VP bundles in all four sections of their research papers, in comparison to the soft fields. As was noted earlier (see Section 5.2.2), NP-VP bundles comprise both noun phrases and verb phrases (fragments or clauses). These bundles are often structurally complete and perform specialized and readily discernable communicative functions (e.g., *are associated with higher levels of, analysis was performed using*). These results corroborate those of Cortes (2004), who also found that science writing made greater use of bundles that incorporated both verb and noun phrase structures (e.g., verb + complement noun/adjective phrase and noun phrase + verb + complement clause fragments). Looking at Appendix B, we find that these bundles

are highly conventionalized academic phrases that are commonly used in the Results section of research articles, particularly those written in the hard sciences.

The results of the second phase of the analysis further revealed fundamental disciplinary variation in the use of PP-bundles across sections, with Introductions and Methods in soft knowledge fields containing substantially more instances of bundles from this structural group, compared to the hard sciences. The top five most frequent PP-based bundles used in Introductions and Methods in the soft knowledge fields are: *on the basis of, in the context of, as well as the, for each of the* and *on the other hand.* As Hyland (2008) argues, the frequent use of PP-based bundles of this sort reflects writers' attempt to situate their arguments within a particular context (e.g., *in the context of, in terms of*) and establish links between elements of their discussion (e.g., *as well as the, on the other hand*). Hyland (2008) observed that academic texts (PhD dissertations, MA/MSc theses, and research papers) written in Social sciences contained substantially more bundles beginning with a prepositional phrase. Similarly, Biber and Gray (2016, p. 116) found that prepositional phrases functioning as adverbials or verb complements (e.g., *regarded <u>as one of the</u>*) and prepositional phrases headed by *for* (e.g., *one explanation <u>for this relationship</u>*) were frequently used in Humanities and Social sciences, respectively (see Appendix B for lexical bundles with similar structures). The results of the present research extend the findings of Hyland (2008) and Biber and Gray (2016) by revealing that such prepositional phrases are more likely to occur in Introductions and Methods in the soft fields than in the hard sciences. Moreover, the second part of the analysis further revealed that the frequency of occurrence of verb-based bundles (e.g., *to account for the*) was significantly higher in Methods in the soft disciplines compared to those in the hard sciences, suggesting that writers in the soft field tend to rely on verb-based phrases when describing the methodology used in their research. However, it is important to note that the overall use of VP-based bundles was considerably

lower than that of PP-based bundles in this section. The frequent use of verbal structures have been shown to reflect an elaborated form of discourse (e.g., Biber, et al, 1999; Biber & Gray, 2016). Gray (2015) found that non-finite completement clauses (*there is a need to arrive at a more*) were more commonly used in Humanities and Social sciences than in the hard sciences, which was taken as a reflection of the elaborated style of writing in the soft fields. As shown in Appendix B, most of the highly frequent VP-based bundles used in Methods in the soft fields incorporate clause fragments (e.g., *are more likely to be, to take into account*). And so, it appears that the presentation of research methods in the soft fields is more discursively elaborate than in the sciences.

The third phase of the quantitative analysis focused on the communicative functions of the identified bundles in the IMRD sections across disciplines. The analysis revealed systematic differences between hard and soft knowledge fields in their reliance on using different types of research-oriented bundles when writing Introductions and Discussions, with writers in the sciences using substantially more bundle types from this category in these sections. This finding is in line with Hyland (2008), Durrant (2017) and Omidian, et al, (2018) who also found a greater concentration of research-oriented bundles in hard science texts. It is worth noting that the majority of these bundles in hard science Introductions and Discussions were found to contribute to the description of the physical aspects of research objects/materials (more on this below). It was also found that the soft disciplines made significantly more use of bundle types and tokens from text- and participant-oriented categories when describing the methodology used in their research. As Hyland (2008) notes, this seems to reflect the more discursive nature of argument in soft disciplines in which persuasive strategies are carried out in a more interpretative, and less empiricist, fashion. This corroborates the patterns of use observed in the second phase of the analysis, which showed

that authors in Humanities and Social sciences made extensive use of bundles from structural groups that reflect more elaboration in describing research methods.

My further investigation focused on the qualitative analysis of the identified bundles in their surrounding context. Specifically, I examined those bundles that were distinctive of each disciplinary field in a certain section. The analyses highlighted a number of key findings. First, the analysis showed that many of the distinctive research-oriented bundles in both Introduction and Discussion sections were employed by science writers to specify the physical attributes of the following noun, describing its tangible features. In the Introduction section, these bundles were found to be used by science writers to provide a detailed explanation of the research processes carried out in previous studies, with the aim of indicating a gap in prior research, emphasizing the centrality of the study being introduced, and explaining the operational nuances of the research activities involved in conducting the reported study. In the Discussion section, research-oriented bundles were employed by these writers to describe procedures and processes, the shape of the material used, and the quantitative aspects of the observations being discussed. The analysis further revealed that a certain kind of research-oriented bundles was distinctively used in Introductions in Humanities and Social sciences. More specifically, it was found that writers in these disciplines used research-oriented bundles as intangible framing devices to specify abstract and intangible characteristics of the following noun (e.g., *the validity of the*). Previous research has shown that the use of such framing devices is the characteristic feature of writing in the soft fields, especially Humanities (e.g., Durrant, 2017; Omidian, Siyanova-Chanturia, & Durrant, 2020). Interestingly, the qualitative analysis showed that about four-fifth of intangible framing bundles were completely absent in hard science Introductions.

Second, the qualitative analysis revealed key disciplinary differences in the ways in which the communicative aims of Methods and Results were fulfilled by authors in hard and

soft disciplines. One interesting example is the use of attitudinal/modality expressions (e.g., *makes it possible to, are more likely to*) by Humanities and Social sciences writers to indicate and frame the possibility of an action or the forthcoming proposition in Methods and Results sections. This suggests that writers in these fields tend to provide tentative support for their choice of a particular approach/methodology, avoid overgeneralizations of their results, and present their findings in a less assured way. Fulfilling the communicative purposes of these sections through the use of such bundles can also be viewed as an indication of the writer's need to anticipate possible objections with regards to the methodology used or the findings obtained. Another key difference is the use of distinctive text-organizers functioning as either transitions between different textual elements (e.g., *on the one hand*) or framing devices that situate arguments within a specific context (e.g., *in terms of the*). These bundles made up about four-fifth of distinctive text-organizers found in the Methods sections written Humanities and Social sciences. In comparison, the limited number of distinctive text-organizers employed by hard science writers in this section were mostly used as structuring signals to direct the reader to specific parts of the text, such as tables and graphs (e.g., *are given in Figure*).

5.5. Conclusion

The intellectual boundaries drawn by disciplinary frameworks can profoundly affect the communication and transmission of knowledge in different academic domains. The findings reported in this study support this view by showing that disciplinary fields considerably vary with respect to the rhetorical strategies they adopt to fulfill the rhetorical aims of the main parts of their research papers. The study also revealed a strong form-function connection between MWEs and the communicative purposes that they perform in these sections. Highlighting the differences that emerge from the ways writers use this form-function

connection to construct knowledge in different academic domains has great potential to reinforce our understanding of disciplinary writing. Studies of this kind can also be of great help to graduate students and early-career researchers who wish to participate in scientific activities in their field but have yet to acquire the disciplinary knowledge required for such an undertaking. Future research can investigate other linguistic devices, such as syntactic features, in multiple sections of research articles and conduct a cross-sectional analysis of these features and their associated communicative functions through the lens of disciplinary specificity.

5.7. Acknowledgements

Chapter 6

Study 3: New dimensions of variation in research writing

6.1. Introduction

Over the past four decades, writing for research publication has been an active area of research in various lines of inquiry within the field of applied linguistics (see Lillis & Curry, 2016 for a comprehensive review). The major reason for this long-standing interest is mainly related to the distinct linguistic characteristics of this particular form of knowledge communication in academia. A prototypical example of such characteristics is the extensive use of nouns, nominal phrases, and embedded phrasal structures, which results in the highly compact and informational discourse of academic research prose (Biber, 1988; Biber & Gray, 2016). Gaining control of this unique prose style requires a strong understanding of the linguistic particularities that characterize research writing and set it apart from other written genres. Developing such a specialized competence, however, usually demands prolonged exposure to the norms and conventions that govern such linguistic differences (see e.g., Cortes, 2004; Romer, 2009). This protracted development often poses major challenges to those who wish to participate in scholarly activities in their fields but have yet to acquire the discourse competencies required for such a feat. These challenges are further compounded by the heterogeneous nature of knowledge production across academic fields. That is, the norms and conventions of research writing can be differentially affected by those of disciplinary communities and their routine scholarly practices. Thus, writing for research publication seems to involve facing the dual task of mastering a particular mode of knowledge communication while gaining familiarity with the linguistic demands associated with disciplinary conventions.

Research in this area has focused on investigating frequently used lexical and grammatical features that distinguish conventional discourse practices in research writing. The rationale behind this research agenda stems from the premise that certain lexical and grammatical devices serve particular communicative functions, which are often characterized by conventional discourse and community-specific practices. This body of research has been guided by two main approaches. One approach has been to investigate the lexico-grammatical markers of a particular communicative function in research writing (for a review of studies adopting this approach see Chapter 2, Section 2.3.4.3). In this deductive research approach, the theoretical framework for investigating lexico-grammatical patterns is determined at the outset of the research and the entire analysis is then conducted on this a priori basis. A different approach altogether has been to use statistical methods to locate systematic co-occurrence patterns of linguistic features and then identify the underlying discourse functions that give rise to such patterns of language use. In this inductive analytical approach, which was developed by Biber (1988), communicative functions underlying co-occurrence patterns of linguistic features emerge from the analysis with no preconceived functional/discourse theory guiding their identification (see for further discussion Chapter 2, Section 2.3.4.3).

While the findings from this body of research have provided important insights into the linguistic characteristics of writing for research publications, they have shared certain key constraints that have made it difficult to arrive at a comprehensive view of this particular genre of writing. First, the majority of the studies on this topic have either focused on the linguistic characteristics of academic research prose as a whole or a particular section of these texts. Specifically, very few studies have considered the impact of intratextuality on systematic patterns of lexico-grammatical features in research writing (e.g., Biber & Finegan, 2001). Fewer still have investigated the influence of this factor on research writing across a

range of academic disciplines. This is surprising because studies have long emphasized the importance of considering possible internal variations within academic genres (e.g., Bhatia, 1993; Swales, 2004; Swales & Feak, 2004). Secondly, research in this area has largely overlooked the linguistic differences that may exist between research articles written by L1 and LX authors. The dominance of English as the language of publishing in contemporary academia has raised the question of whether L1-English authors have a head start in acquiring the literacy skills associated with writing for research publication purposes (see Flowerdew, 2013a, 2013b). While claims and anecdotal evidence about L1-LX differences in expert research writing abound, little comparative linguistic data have been offered in evidence (see Hyland, 2016).

Acknowledging these gaps, the present research sought to investigate the linguistic characteristics of research writing through the lens of lexico-grammatical features that characterize the communicative purposes of main sections of empirical research articles published by L1- and LX-English writers from across different academic disciplines. In doing so, the multidimensional research approach developed by Biber (1988) was adopted through which the functional dimensions associated with systematic co-occurrence patterns of a wide range of lexico-grammatical features were identified. These dimensions were then used to provide a functional characterization of research writing and highlight possible linguistic differences arising as a result of factors such as disciplinarity, intratextuality, and possible influences of L1-versus LX-English expert writing.

6.2. Methodology

6.2.1. Conducting a full multidimensional analysis

6.2.1.1. Lexico-grammatical annotation of the DCRA

As a first step in conducting a full MD analysis (see Chapter 2, Section 2.3.4.3.), all texts in the DCRA were annotated for various lexical and grammatical features using the Biber tagger (see Biber, 1988, Biber et al, 1999). The Biber tagger employs on-line dictionaries, probability-based algorithms, and context-based rules to assign linguistic information (i.e., tags) to words based on their semantic and syntactic properties. In doing so, the tagger uses five 'tagfields' to analyze and identify semantic and syntactic characteristics of lexical items. In the first tagfield, the tagger identifies the primary grammatical (part of speech) category of words. In the remaining tagfields, additional syntactic and/or semantic information is identified for words or syntactic structures. For instance, participial attributive adjectives (e.g., _sleeping dogs_) are identified as adjectives by the tag _jj_ in the first field, as attributive by the tag _atrb_ in the second field, and as being participial by the tag _xvbg_ in the fourth field (the final tag sequence assigned to the participial attributive adjective _sleeping_ is _jj+atrb++xvbg+_). In another example, the tag sequence _whp+rel+subj++_ is used to characterize the relativizer _which_ as a 'WH pronoun' with subject gaps (e.g., _a pattern of result which is in line with_).

The Biber tagger is designed to identify a wide range of linguistic features in spoken and written corpora (for further information about the tagger see Biber 1988, p. 211; Biber et al., 1999, p. 35; Biber, 2006, p. 28). Using the tagger and its companion 'TagCount' program (more on this below), over one hundred and forty lexical and grammatical features were identified for all texts in the DCRA.

6.2.1.2. Quantifying lexico-grammatical features

In the second step of the analysis, the rates of occurrence of the tagged linguistic features in the corpus were quantified using the associated TagCount program (see Biber 1988, p. 75). The TagCount program relies on both simple and rule-based tag searches to generate frequency data for various lexical and grammatical features. For example, while the frequency of all public verbs (e.g., *assert, complain*) in the corpus is calculated by tallying the rates of occurrence of the tag *vb+vpub+++*, determining the frequency of *that*-complement clauses controlled by stance verbs requires the program to employ rule-based algorithms that combine lexical and tag information to identify stance verbs (e.g., *appear, seem* marked by the tag *vb+seem+++*) followed by the word *that* tagged as a complementizer (i.e., *tht+vcmp+++*), and not a relativizer (e.g., *tht+rel+++*). It should be noted that the program provides frequency counts for the tagged features at different levels of granularity. For instance, in addition to providing the total rate of occurrence of *that*-complement clauses controlled by all stance adjectives (e.g., *it is possible that the*), the program also provides the individual frequency counts of *that*-complement clauses controlled by specific stance adjectives (e.g., *that*-complement clauses controlled by attitudinal, factive, and likely adjectives, *it is clear that*). Therefore, there are hierarchical dependencies among the features provided by the program (more on this below). The frequency counts of all lexico-grammatical features are then normalized (per 1,000 words) and recorded for each text in the corpus. In the case of the DCRA, this procedure resulted in a dataset comprising 349,280 data points for a hundred and forty-eight linguistic features.

6.2.1.3. Selection of the linguistic variables

In the third step of the analysis, the linguistic features that were suitable for inclusion in the final analysis were determined. Determining which linguistic variables to include in a full

MD analysis is extremely important, as the success of the statistical procedure and the underlying functional dimensions that emerge from the analysis are contingent on the final set of features selected (Biber, 1988, p. 71; Egbert & Staples, 2019).

As was mentioned above, many of the linguistic features provided by the TagCount program are not independent of each other. In addition, the output of the program also includes features that are highly overlapping and measure the same linguistic construct. For instance, the program reports the frequency of both public and communication verbs, although they are almost the same set of verbs. The reason for this is that the TagCount program is based on two versions of the Biber tagger. For example, while the lexical class 'public verbs' is based on the version of the tagger used in Biber (1988), the feature 'communication verbs' is based on the version developed by Biber et al. (1999). The inclusion of overlapping and hierarchically dependent features (e.g., 1st, 2nd, 3rd person pronouns vs all pronouns) would result in the problems of multicollinearity (highly correlated variables, e.g., above .90) and singularity (redundant variables) in the final factor analysis (see Tabachnick & Fidell, 2007, p. 88). Therefore, care should be taken to select conceptually distinct features for the final analysis. The goal here is to include the widest possible range of linguistic features, while maintaining statistical precision (see, e.g., Biber, 1988, p. 72; Conrad & Biber, 2001, p.15).

Principled decisions were made to avoid overlapping features and include those that could explain the greatest proportion of variance in the corpus. These decisions were based on the results of various pilot factor analyses. The selection procedure began with features defined at most specific level. Various pilot factor analyses were then run to determine which features were likely to contribute statistically meaningful information to the explanation of variance among texts. Based on the results of these preliminary factor analysis runs, variables with low communalities (reflecting low shared variance with the overall factorial structure)

were eliminated (see Tabachnick & Fidell, 2007, p. 621). In addition, redundancies and overlapping features were avoided by combining variables and removing those with low communalities, low frequency, and low factor loadings[3]. For instance, in the case of hierarchically dependent variables, semantic subclasses of features (e.g., prediction modals, necessity modals, and possibility modals) were retained in the analysis if they did not show low communalities or factor loadings; otherwise, they were replaced by more general, superordinate features (e.g., all modal verbs). This procedure resulted in a set of 50 linguistic features for the final factor analysis (see Appendix C, Table C1).

6.2.1.4. Final factor analysis

As the fourth step in the procedure, the final factor analysis was conducted with the selected features using Principal Axis Factoring in SPSS (version 26.0). A four-factor solution was selected as optimal. This decision was based on the inspection of the scree plot of eigenvalues which showed a clear break between factors four and five (see Appendix C, Figure C1). The four factors accounted for 40.1 per cent of the cumulative shared variance[4]. The factor solution was rotated using a Promax rotation, which permits minor correlations among the extracted factors (see Biber, 1998, p. 85). The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy showed a high degree of factorability of the selected features (KMO = 0.812, *meritorious*; see Kaiser, 1974), which surpassed the minimum requirement of 0.60 (see Tabachnick & Fidell, 2007, p. 614). Barlett's test of sphericity, as an additional measure of factorability, was also significant (Approximate Chi-Square = 34236.219, df = 1225, p = 0.000), indicating that the null hypothesis that the correlations in the correlation

---

[3] Factor loadings, or weights, represent the relationship (correlation coefficients) between each variable and the underlying factor.

[4] The amount of variation explained by the factors is similar to those of previous MD studies (see Egbert & Staples, 2019, for a survey of previous MD studies).

matrix are zero can be rejected (Tabachnick & Fidell, 2007). Both these measures indicated that conducting factor analysis was suitable for the selected features.

6.2.1.5. Factor scores

In the final step of the analysis, factor scores were computed for each text in the corpus. In doing so, standardized scores (z-scores) were calculated for linguistic features with salient factor loadings (greater than ±0.30, see Egbert & Staples, 2019 for the cut-off criteria used in previous MD studies). The standardization procedure served to preclude the possibility of common features (e.g., prepositions) having an inordinate influence on factor scores (Biber, 1988, p. 94). Following this, factors scores were calculated for each text by subtracting the sum of the standardized scores of features loaded on the negative pole of each factor from the sum of the standardized scores of features loaded on the positive pole of that factor. In cases where a feature had salient loadings on multiple factors, it was included in the factor score of the factor on which it had the highest loading (also see Biber, 1988, p. 93). Once factor scores of texts in the corpus were computed for all the four extracted factors, the mean values of the factor scores were calculated for various sub-corpora of the DCRA. General linear models (in SAS) were then used to compare various register groupings of texts in the corpus.

The following section presents the results of the MD analysis conducted in this study and discusses how the functional construct, or 'dimension', underlying each of the extracted factors were interpreted through a detailed functional analysis of co-occurring features.

6.3. Results and discussion

6.3.1. Dimensions of variation in expert research writing

Table 6.1. presents the lexico-grammatical features defining three primary dimensions of

variation in research writing[5].

Table 6.1. *Summary of the important linguistic features (with loadings larger than +/-.30) on each dimension*

Dimension 1
Features with positive loadings:
predicative adjectives (.51), general adverbs (.50), pronouns: demonstrative (.46), pronouns: *it* (.56), verb *BE* (.62), present tense verbs (.65), necessity/obligation modals (.40), possibility modals (.58), prediction modals (.46), *that*-clauses: controlled by stance adjectives (.41), *that*-clauses: controlled by stance verbs (.44), *to*-clauses: controlled by stance verbs (.37), adverbial conjunctions (.43), epistemic adjectives in predicative position (.34), attitudinal adjectives in predicative position (.32), attitudinal adjectives in other contexts (.33), stance adverbials (.32), conditional subordinating conjunctions (.31)

Features with negative loadings:
common nouns (-.40), nouns: pre-nominal modifiers (-.40), [past tense verbs (-.38)], [word length (-.32)]

Dimension 2
Features with positive loadings:
common nouns: abstract/process (.37), common nouns: cognitive (.39), nominalization (.67), pronouns: 3rd person (.42), verbs: communication (.42), word length (.80), type-token ratio (.43), common nouns: human (.33), verbs: suasive (.33), phrasal coordinating conjunctions (.32); *to*-clauses: controlled by stance nouns (.31)
Features with negative loadings:
common nouns: concrete/tangible (-.35), common nouns: quantity (-.42), prepositions (-.38), [predicative adjectives (-.34)]

Dimension 3
Features with positive loadings:
past tense verbs (.54), verbs: activity (.48), passives: agentless (.42), adverbials: time (.38), verbs: aspectual (.35), present progressive verbs (.32),
[verbs: communication (.39)], [pronoun: 3rd person (.38)]
Features with negative loadings:
attributive adjectives (-.44), [word length (-.32)]

---

[5] While the factorial structure of Dimension 4 was relatively strong, it was not readily interpretable (see Appendix C, Table C1). Thus, this dimension was not considered further in the present study.

Detailed considerations of these features and their shared communicative functions provided the basis for the functional interpretation of the extracted factors as the following linguistic dimensions:

Dimension 1: Evaluative and elaborated versus informational discourse

Dimension 2: Conceptual and abstract versus concrete language

Dimension 3: Past-time procedural discourse

In the interpretation of a factor as a linguistic dimension, the communicative functions shared by the identified features are sought to explain the functional dimension underlying the factor in question (Biber, 1988, p. 91). This is based on the premise that the communicative demands of texts (academic or non-academic) have the potential to give rise to systematic co-occurrence patterns of certain lexico-grammatical features (see, e.g., Biber, Egber, & Keller, 2020). That is, co-occurrence patterns among linguistic devices are highly unlikely to be coincidental. Rather, they appear to be inextricably linked to communicative purposes and situational characteristics of texts.

The following sections discuss the details of these functional interpretations and the relations among texts with respect to each dimension.

6.3.2. Dimension 1: Evaluative and elaborated vs informational discourse

6.3.2.1. Interpretation of Dimension 1

Table 6.1 above shows that Dimension 1 comprises a wide range of linguistic features on the positive end of the factor and a few features with negative weights. The negative pole of Dimension 1 is very similar to those of previous MD studies (e.g., Biber, 1988; Biber, 2006a; Gray, 2015; Egbert, 2015). Four features loaded on the negative end of Dimension 1 (common nouns, pre-modifying nouns, past tense verbs, word length), two of which had larger weights on other factors (past tense verbs, word length; presented in [brackets] in

Table 6.1). The co-occurrence of these features has been shown to reflect a high informational focus and the density of information in a text (see, e.g., Biber, 1988; Biber, 2006a). In expert research writing, nominal features, such as common nouns and pre-modifying nouns, afford the writer the opportunity to communicate a wealth of information by condensing it into few words (see, e.g., Biber & Gray, 2016). Common nouns provide this opportunity by encompassing meanings from a wide and diverse array of semantic domains. Biber (2006a) classifies common nouns based on their meanings into eight major semantic categories: animate/human (*manager*), concrete (*cylinder*), technical (*ray*), cognitive (*view*), place (*middle*), quantity (*frequency*), group (*university*), process/abstract (*function*) nouns. Because they can fulfil a variety of semantic roles, common nouns are often the primary bearers of meaning and information in academic discourse (e.g., Biber & Gray, 2016, p. 114). Thus, the high frequency of common nouns in a text can be associated with communicative purposes that demand a high informational focus. Dimension 1 shows that these nouns are also often pre-modified (or qualified) by other nouns, resulting in multi-word combinations comprising two or more nouns (e.g., *food safety management programs*). The concatenation of nouns in this way enables authors to add additional layers of information to the meaning of the head noun with fewest words possible. Frequent use of such structural compressions often results in inexplicit expression of meaning and a dense packaging of information (see discussion in Biber & Gray, 2016, Chapter: 6). This process also directly contributes to the creation of technical terminologies and domain-specific meanings, which play a key role in the transmission of knowledge across academic disciplines. The following excerpt illustrates the use of these features in a text with a high negative score on Dimension 1:

1) CHEMISTRY: *METHODS*

The burner *diameter* was 50 mm surrounded by a 5 mm annular ring, also of sintered metal, providing an inert shield gas *flow*. A conical chimney with a 45-degree *taper*,

whose base matched the <u>burner</u> *diameter* fed a <u>steel</u> *tube*. At the top of the chimney, a circular molybdenum ring held the mesh-supported catalyst within the <u>post-flame</u> *gases*. The <u>flame equivalence</u> *ratio* was varied, adjusting the <u>fuel flow</u> *rate* to the burner while maintaining the <u>air</u> *flow* at 115 SLPM. The <u>flame gas</u> *composition* was calculated with a chemical <u>equilibrium</u> *code* on the basis of the measured <u>post-flame</u> *temperature*, initial <u>fuel</u> *identity*, and the <u>fuel-air equivalence</u> *ratio* as inputs. Table 1 lists the fuels and <u>fuel/air equivalence</u> *ratios* for the flames studied here and the corresponding <u>equilibrium gas</u> *compositions*

The primary positive features on Dimension 1 include modifiers (predicative adjectives, general adverbs), pronouns (pronoun 'it', demonstrative pronoun), verb *be*, present tense verbs, *that*-clauses controlled by stance adjectives*, that*-and *to*-clauses controlled by stance verbs, modals (possibility modals, necessity/obligation modals, prediction modals), and adverbial conjunctions. The majority of these features function together to provide a structure for conveying evaluations and stance meanings. For example, predicative adjectives tend to complement a copular *be* to present the author's epistemic/attitudinal stance on the nominal element in subject position (e.g., *the effect on antagonistic behavior <u>is particularly strong</u>*) (see discussion in Biber et al, 1999, Section 5.5). These adjectives can also control complement clauses: *to*-clauses (e.g., *such statements are more <u>likely to</u> shape*) or *that*-clauses (e.g., *it is <u>conceivable that</u> protein abduction*). Both of these structures provide a frame for the expression of stance meanings (Biber, 2006b). Dimension 1 shows that texts in the corpus systematically varied with respect to the use of *that*-clauses controlled by stance adjectives. In academic writing, adjectives controlling *that*-clauses typically occur in extraposed constructions, where pronoun 'it' is the grammatical subject (see Gray & Biber, 2015). Such structures allow writers to foreground their evaluative stance on the information presented in the following clause (for further discussion see Biber,

et al, 1999, Section 7.5; Biber, 2006a, Chapter 5). Excerpt 2-3 illustrates the co-occurrence patterns of these linguistic features:

2) DENTISTRY: *DISCUSSION*

Despite this limitation, the power calculation was based on PD as the primary outcome of interest and we <u>are confident that</u> clinical comparisons between diabetes and non-diabetes subjects <u>are valid</u> whereas observations resulting from the subgroups analyses <u>are less robust</u>

3) SOCIOLOGY: *DISCUSSION*

<u>It is likely that</u> the possibility of utilising these factors varies among young adults. For instance […]. However, <u>it is possible that</u> the results presented here could have been different if other poverty measures had been used

Other clausal features with salient positive loadings on Dimension 1 are *that*-and *to*-clauses controlled by stance verbs (e.g., *these constraints <u>tend to</u>*, *they <u>contend that</u> anxiety*). Depending on the semantic domain of the controlling verb, these structures can convey various evaluative and stance meanings, including the author's epistemic judgments, their views and arguments, and the amount of confidence they place in a proposition (see Biber 2006b; Charles, 2006; Hyland, 2002a; Hyland & Tse, 2005). Similar to adjectival predicates, these verbal predicates can also take extraposed *that*-clauses (e.g., *it seems that*).

The positive end of Dimension 1 further indicates that texts in the corpus varied in the expression of stance through the use of modal verbs: possibility modals, prediction modals; necessity/obligation modals. Modals of possibility are used to indicate the possibility of an action (e.g., *quantity of stereotypes <u>could be used as</u>*) or to convey the author's epistemic perspectives in a less direct and assured manner (e.g., *assuming that students do perceive such support <u>may be a mistake</u>*). Modals of prediction are employed to make (hypothetical) predictions that are often based on evaluations/judgments (e.g., *more complex designs <u>will</u>*

essentially increase). Modals of necessity provide a frame to express the necessity or obligation of the forthcoming proposition (e.g., *scientific political ideas should be considered within*) (for further discussion of modality see Biber et al, 1999, p. 483).

Finally, adverbial conjunctions also co-occurred meaningfully with other features on the positive end of Dimension 1. Adverbial conjunctions serve to connect stretches of discourse, creating lexical cohesion and ensuring the logical of flow of discussion (e.g., *however, therefore*). The statistical co-occurrence of this feature with the above stance devices on Dimension 1 seems to reflect the evaluative and discursive patterns of argument in texts with high scores on this dimension. Excerpt 4 illustrates these co-occurrence patterns in a text with a high positive score on Dimension 1:

4) MANAGEMENT: *DISCUSSION*

Given the time savings involved, organizations <u>might</u> <u>be</u> <u>able</u> to interview more candidates by using a nonverbal medium. <u>It</u> <u>is</u> even <u>possible that</u> a written BDI <u>could</u> <u>be</u> used for e-recruitment, and then high-scoring candidates <u>could</u> <u>be</u> processed further […]. One potential consideration, <u>however,</u> is that the higher correlation with cognitive ability <u>may</u> make direct assessment of cognitive ability less <u>viable</u>. <u>Regardless,</u> included in such research <u>should</u> <u>be</u> an assessment of whether […]. <u>However,</u> if assessment of more maximal capabilities <u>is</u> the goal then a traditional BDI <u>would</u> not <u>seem to</u> <u>be</u> a good fit and a variation where interview constraints <u>are</u> <u>relaxed</u> <u>might</u> <u>be</u> more <u>effective</u>. […] We <u>believe that</u> finding effective intermediate points <u>is</u> <u>very</u> <u>possible</u>

Excerpt 4 reflects the author's attempt at structuring and presenting ideas and evaluations, as convincing and coherently organized arguments, using adverbial conjunctions and various stance and evaluative devices. The patterns of argument and reasoning in the above text sample can also be considered interactive, seeking to convince readers or turn their attention

to the necessity of a proposition through the logical force of the argument. Clearly, the linguistic manifestations of this writer-reader interaction are not overtly interpersonal (cf. Biber, 1988, p. 105); Rather, they seem to conform to what is often referred to as 'academic interactions' (see Hyland, 2001, 2002b, 2004).

In sum, the patterns of language use shown in Excerpt 4 appear to be fundamentally different from those shown in Except 1. Considering both positive and negative features, Dimension 1 seems to distinguish communicative situations that require evaluations and elaborations from discourses with a high informational focus.

6.3.2.2. Register variation along Dimension 1

The results of an ANOVA (from the GLM procedure in SAS) showed that Dimension 1 was a significant and strong descriptor of register variation in expert research writing ($F = 129.3$, $p = .<0001$, $R^2 = .43$)[6]. This variation is summarized in Figure 6.1. The figure shows that Discussions written in the soft fields have very high mean scores on dimension 1, while Methods in the sciences are marked by low mean scores. This indicates that writing Discussions in soft knowledge fields is mostly associated with stance taking, evaluation, and elaboration, whereas the focus of the Methods section in the sciences is primarily informational. However, this dichotomy merely scratches the surface of the variation patterns revealed by Dimension 1. On closer inspection, it turns out that there are fundamental differences among IMRD sections with respect to the use of features loaded on Dimension 1, that these features are used to different degrees in hard and soft knowledge fields across the sections, and that there are subtle, yet important, differences between L1 and LX expert writing across the four sections. Figure 6.2 plots these differences.

---

[6] The R^2 values reported for the factors is a measure of their effect sizes, which describe the proportion of variance in factor scores that is associated with register categories (see Tabachnick & Fidell, 2007, p. 54).

EVALUATIVE AND ELABORATED

12	L1 Soft Discussion ($M = 12.56$)
11	
10	L2 Soft Discussion ($M = 10.77$)
9	
8	
7	
6	L1 Hard Discussion ($M = 6.53$)
5	
4	
3	
2	L1 Soft Introduction ($M = 2.32$)
	L2 Soft Introduction ($M = 2.30$)
	L2 Hard Discussion ($M = 2.11$)
	L1 Soft Results ($M = 2.08$)
1	L2 Soft Results ($M = 1.98$)
0	
-1	
-2	L1 Hard Introduction ($M = -2.22$)
	L1 Soft Methods ($M = -2.28$)
	L2 Soft Methods ($M = -2.59$)
	L2 Hard Introduction ($M = -2.66$)
-3	L1 Hard Results ($M = -3.60$)
-4	
-5	L2 Hard Results ($M = -5.52$)
-6	
-7	
-8	
-9	
-10	L1 Hard Methods ($M = -10.33$)
-11	
-12	L2 Hard Methods ($M = -12.37$)

INFORMATIONAL

Figure 6.1. Dimension 1 mean scores for the IMRD sections written by L1 and LX authors across hard and soft domains ($F = 129.3$, $p = .<0001$, $R^2 = .43$)

149

EVALUATIVE AND ELABORATED

11	Soft Discussion (*M = 11.69*)	
10		
9		L1 Discussion (*M = 9.60*)
8	Discussion (*M = 8.05*)	
7		
6		L2 Discussion (*M = 6.51*)
5		
4	Hard Discussion (*M = 4.32*)	
3		
2	Soft Results (*M = 2.35*) Soft Introduction (*M = 2.31*)	
1		
		L1 Introduction (*M = 0.08*)
0	Introduction (*M = -0.02*)	L2.Introduction (*M = -0.14*) L1 Results (*M = -0.71*)
-1	Results (*M = -1.20*)	L2 Results (*M = - 1.70*)
-2	Hard Introduction (*M = -2.44*) Soft Methods (*M = -2.44*)	
-3		
-4	Hard Results (*M = -4.56*)	
-5		
-6	Methods (*M = -6.82*)	L1 Methods (*M = -6.24*)
-7		L2 Methods (*M = -7.40*)
-8		
-9		
-10		
-11	Hard Methods (*M = -11.35*)	

INFORMATIONAL

Figure 6.2. Dimension 1 mean scores for different register categories

The distribution of IMRD sections along dimension 1 show a dichotomous relationship

between Methods and Discussion sections (see the left-hand side of Figure 6.2). In

comparison, Introduction and Results have relatively similar scores on this dimension. (These

relations are further supported by the post-hoc comparisons reported in Appendix C, Table

C2) In general, it seems that the communicative purposes of the Discussion section demand a

reliance on authorial stance, evaluation and elaboration, while the discourse of Methods is extremely impersonal and primarily focused on communicating factual information.

Continuing towards the middle of the plot, we notice that these patterns are affected by disciplinary specificity. Most noticeable are the patterns of use for Introductions and Results, indicating that these sections in the soft fields are substantially more evaluative and elaborated compared to the hard sciences. Figure 6.2 also shows that discussion of findings is the only part of the research article in which authors in the hard fields tend to heavily rely on evaluation and elaboration features. Post-hoc analysis showed that soft knowledge fields had significantly higher scores on Dimension 1 than the sciences across all the four sections, with the magnitude of these differences being the largest in Methods and Discussions (see Appendix C, Table C3).

Finally, the right-hand side of Figure 6.2 shows that the use of Dimension 1 features by L1 and LX expert writers is relatively comparable across all the sections except for Discussions. Interestingly, the patterns of variation revealed by Dimension 1 show that, compared to LX authors, L1 writers made significantly greater use of the positive Dimension 1 features when discussing the findings of their study (see Appendix C, Table C4), suggesting a higher degree of evaluation and elaboration in the discussions written by these authors.

6.3.3. Dimension 2: Conceptual and abstract vs concrete language

6.3.3.1. Interpretation of Dimension 2

Dimension 2 consists of 11 positive and four negative features (Table 1). The primary positive features on this dimension include two classes of common nouns (cognitive and abstract/process nouns), nominalization, 3rd person pronouns, communication verbs, type/token ratio, and word length. Cognitive nouns are employed to convey meanings related

to cognitive entities (e.g., *perspective, judgement*), typically referring to abstract concepts associated with human agents (see Gray, 2016, p. 91). Abstract/process nouns are used to refer to intangible processes and concepts (e.g., *function, argument*). Abstract nouns are also created through nominalization, a process through which meanings are decoupled from their verbal/adjectival realizations and reconstructed using nominal forms which often have no perceptual correlate (e.g., *commitment, responsibility*) (see, Biber & Gray, 2016; Halliday, 2004 for further discussion). The co-occurrence of these nominal elements thus seems to signal a form of discourse primarily concerned with descriptions of concepts and abstract entities:

5) APPLIED LINGUISTICS: *INTRODUCTION*

A historical <u>look</u> at the <u>evolution</u> of <u>research</u> on the <u>notion</u> of teacher <u>thinking</u> in mainstream <u>education</u> reveals that prior to the 1970s, this <u>line</u> of <u>enquiry</u> was primarily guided by the <u>process-product</u> <u>conceptualization</u> of teaching. Couched within this <u>paradigm</u>, teaching was primarily viewed in the <u>light</u> of the <u>outcomes</u> it produced.

6) MANAGEMENT: *DISCUSSION*

We suggest that, as mature <u>knowledge</u> is being incorporated in an increasing number of <u>innovations</u>, they lose their <u>novelty</u>. Therefore, inventors are better off leveraging that <u>knowledge</u> before […]. This entails monitoring the <u>pace</u> of <u>knowledge</u> <u>adoption</u> in the industry. <u>Consideration</u> of the distance and <u>adoption</u> <u>contingencies</u> can

Arguments in the above text samples are conveyed through the author's careful selection of abstract nouns that are very specific in meaning. Dimension 2 shows that texts with high frequency of use of these nouns are also marked by diverse lexical choices and long words (as indicated by type/token ratio and word length). A high degree of lexical diversity reflects a precise and careful selection of vocabulary (see discussion in Jarvis, 2013). The frequent

use of long words also implies the reliance of these texts on specific meanings: longer words are often more precise in meaning than shorter words (Zipf, 1949). The co-occurrence of these features indicates that texts with high positive scores on Dimension 2 are characterized by precise choices of value-laden words used to describe conceptual and abstract entities.

Communication verbs also have a salient positive loading on Dimension 2. Academic verbs of communication are typically used to discuss and communicate abstract concepts such as arguments or claims put forth by individuals (e.g., *argue*, *discuss*, *suggest, state*) (see Biber, 2006b, Hyland, 2001). The co-occurrence of these lexical verbs with the abstract nouns discussed above can be seen as representing discoursal practices concerned with developing conceptual understanding of abstract entities and arguments. These co-occurrence patterns are illustrated in the following excerpt from a text with a high score on Dimension 2:

7) BUSINESS: *INTRODUCTION*

The <u>pursuit</u> of entrepreneurial <u>opportunities</u> is at the <u>heart</u> of <u>entrepreneurship</u>. While there are many <u>activities</u> associated with <u>opportunity</u> <u>pursuit</u>, <u>judgments</u> and <u>decisions</u> have been documented as fundamental to the <u>process</u>. […] Cognitive <u>science</u> <u>research</u> indicates that one's past <u>inaction</u> <u>decisions</u> can produce downstream <u>effects</u> when it comes to the <u>judgments</u> and <u>decisions</u> that follow. In <u>line</u> with this, anecdotal <u>evidence</u> <u>suggests</u> such <u>dynamics</u> in entrepreneurs' <u>decision making</u>, as well-regarded entrepreneurs like Amazon founder Jeff Bezos frequently <u>discuss</u> past <u>inaction</u> <u>decisions</u> as a <u>factor</u> in how they think about potential <u>opportunities</u> today.

In contrast, the negative end of Dimension 2 highlights a form of discourse that rarely involves descriptions of abstract theoretical concepts, and instead relies on explanation of concrete and demonstrable evidence through frequent use of concrete and quantity nouns (e.g., *cylinder*, *temperature*), which are also often expanded upon through prepositional

phrases (e.g., *rays at different angles*). The following excerpt illustrates the discourse of texts with high negative scores on Dimension 2:

> 8) DENTISTRY: *RESULTS*
>
> The control resin and Dycal showed increasing trends in weight loss over time. No significant differences were found in the weight loss of the experimental material amongst all end-points, whereas MTA had a weight increase (percentage weight variation 0) over time. […] In addition, the pH values of the soaking water of the experimental material and MTA did not differ significantly at any time-point except at day 1.

The negative pole of Dimension 2 thus appears to highlight communicative practices that focus on concrete presentation of information, placing emphasis on describing particularities of research objects rather than explicating theoretical concepts.

6.3.3.2. Register variation along Dimension 2

Figures 6.3-6.4 show that the functional construct underlying Dimension 2 is strongly correlated with how disciplines present scholarly knowledge in sections of research articles ($F = 183.5, p = .<0001, R^2 = .51$). At one level, the patterns of variation revealed by Dimension 2 point to some differences across IMRD sections. It can be seen from Figures 6.3-6.4 that Introduction and Discussion sections have higher scores than Methods and Results. Given the strong focus of these sections on explaining and establishing abstract concepts (see e.g., Swales, 1990, 2004), it might be expected that Introductions and Discussions frequently use the features grouped on the positive pole of Dimension 2. However, it turns out that these distributions are strongly affected by disciplinary preferences. Figures 6.3-6.4 show that the IMRD sections with positive scores on Dimension 2 are all written in the soft fields, suggesting that writing these sections in Social Sciences

and Humanities is characterized by patterns of argument that primarily focus on describing

conceptual and abstract entities. In contrast, all sections with negative scores on Dimension 2

belong to the hard sciences, indicating inherent variation in the language of these sections

across hard and soft knowledge fields. Post-hoc analyses showed that hard and soft fields

significantly differed in their use of Dimension 2 features in each of the four sections (see

Appendix C, Table C5).

With respect to possible variation between L1 and LX expert writing, the emerged

patterns showed no systematic differences between the two groups of writers in their use of

the features grouped on Dimension 2. This is further supported by the results of post-hoc

comparisons reported in Appendix C (Table C6).

CONCEPTUAL AND ABSTRACT

6	L1 Soft Introduction (*M = 6.58*)
5	L2 Soft Introduction (*M = 5.83*) L1 Soft Discussion (*M = 5.59*) L2 Soft Discussion (*M = 5.02*)
4	
3	
2	L1 Soft Methods (*M = 2.54*) L2 Soft Methods (*M = 2.25*) L1 Soft Results (*M = 2.15*)
1	L2 Soft Results (*M = 1.77*)
0	L1 Hard Introduction (*M = -0.27*) L2 Hard Introduction (*M = -0.81*)
-1	
-2	L1 Hard Discussion (*M = -2.16*)
-3	L2 Hard Discussion (*M = -3.72*)
-4	L1 Hard Methods (*M = -4.62*)
-5	L2 Hard Methods (*M = -5.44*)
-6	L1 Hard Results (*M = -6.81*)
-7	L2 Hard Results (*M = -7.96*)

CONCRETE

Figure 6.3. Dimension 2 mean scores for the IMRD sections written by L1 and LX authors across hard and soft domains (F = 183.5, p = .<0001, R^2 = .51)

CONCEPTUAL AND ABSTRACT

```
 6 |                                Soft Introduction (M = 6.12)
   |
 5 |                                Soft Discussion (M = 5.30)
   |
 4 |
   |
 3 |                                                                L1 Introduction (M = 3.21)
   |
 2 |  Introduction (M = 2.89)       Soft Methods (M = 2.40)         L2 Introduction (M = 2.57)
   |
 1 |  Discussion (M = 1.25)         Soft Results (M = 1.46)         L1 Discussion (M = 1.12)
   |                                                                L2 Discussion (M = 0.72)
 0 |_____
   |                                Hard Introduction (M = -0.54)   L1 Methods (M = -0.97)
-1 |  Methods (M = -1.25)                                           L2 Methods (M = -1.53)
   |
-2 |  Results (M = -2.88)           Hard Discussion (M = -2.94)     L1 Results (M = -2.25)
   |
-3 |                                                                L2 Results (M = -3.52)
   |
-4 |
   |
-5 |                                Hard Methods (M = -5.03)
   |
-6 |
   |
-7 |                                Hard Results (M = -7.38)
   |
```

CONCRETE

Figure 6.4. Dimension 2 mean scores for different register categories

6.3.4. Dimension 3: Past-time procedural discourse

6.3.4.1. Interpretation of Dimension 3

The positive end of Dimension 3 is characterized by the use of eight features, two of which were found to have larger weights on other dimensions (communication verbs and 3rd person pronouns). The negative end of this dimension consists of only two features (attributive adjectives and word length), one of which loaded with a larger weight on another dimension (word length). And thus, the negative pole of Dimension 3 is primarily characterized by the relative absence of the positive features on this dimension.

The primary positive features on Dimension 3 include agentless passives, two classes of lexical verbs (activity and aspectual verbs), time adverbials, and past tense verbs.

Agentless passives are used to give prominence to the grammatical patient of the verb and minimize the role of the agent (e.g., *samples were immobilized onto*). The co-occurrence of agentless passives with activity verbs (e.g., *combine*, *apply*) seems to reflect communicative practices that center on explaining physical activities and procedures, rather than focusing on the performer of the actions. Aspectual verbs (e.g., *start, complete*) are commonly used to describe stages of progress and sequencing of a given activity or event (see Biber, et al. 1999, pp. 364, 372). Time adverbials are used for temporal reference, marking the timing of events, actions and processes (e.g., *the data was <u>then</u> imported inside a*). Finally, the frequent use of past tense verbs is attributed to discoursal practices that incorporate narrations of past events and happenings. The co-occurrence of these features is illustrated in the following excerpt from a text with a high positive score on Dimension 3:

CHEMISTRY: *METHOD*

Assays were <u>started</u> by the addition of the alcohol oxidase solution and <u>then</u> <u>incubated</u> in a water bath at 30 C. After 10 min, 200 L of a freshly prepared 5 mg/mL Purpald solution in 05 N NaOH <u>was added</u>, and the samples <u>were</u> vigorously <u>vortexed</u> to ensure oxygenation. After an additional 30 min at 30 C, the samples <u>were removed</u> from the water bath and 06 mL of HO <u>was added</u> for a final volume of 10 mL. Absorbance at 550 nm <u>was then determined</u>. Methanol concentrations <u>were</u> <u>determined</u> from a standard curve between 0 and 50 nmol of methanol. Thin slices (1-2 mm) of green bean pods and tomato pericarp <u>were prepared</u> using a sharp knife.

Taken together, it appears that the co-occurrence patterns among the positive features on Dimension 3 mark a form of discourse concerned with the narration of procedures, analytical steps, and specific research-related activities, conveying aspects of methodology and sequences of events with the implied performer of the actions being the researcher.

158

6.3.4.2. Register variation along Dimension 3

Figures 6.5-6.6 show that Methods written in the sciences are characterized by their highly procedural discourse. Methods in Social sciences and Humanities also have moderately high scores on this dimension, indicating that writing these sections in the soft fields relies on procedural discourse features, but to a lesser extent compared to the sciences (see Figure 6.6). The results of post-hoc comparisons showed that Methods sections in the sciences had a significantly higher Dimension 3 mean scores compared to those in the soft fields (see Appendix C, Table C7). This distribution probably reflects a particular emphasis on providing a detailed and accurate description of procedures and analytical steps in the sciences.

Results sections have intermediate values on Dimension 3, suggesting that writing these sections may also rely on some of the positive features on this dimension (e.g., past-tense verbs). At the negative extreme, Introductions in both hard and soft fields are characterized by the near total absence of positive Dimension 3 features (see Figure 6.6), suggesting that communicative practices in these sections rarely involves procedural narrations. Discussions in both domains also have moderately low scores on this dimension.

PAST-TIME PROCEDURAL DISCOURSE

```
5 │ L2 Hard Methods (M = 5.95)
  │ L1 Hard Methods (M = 5.80)
  │
4 │
  │
3 │
  │
2 │ L1 Soft Methods (M = 2.62)
  │
1 │ L2 Soft Methods (M = 1.78)
  │
  │ L1 Hard Results (M = 0.41)
  │ L1 Soft Results (M = 0.20)
0 ├──────────────────────────────────────────────
  │ L2 Hard Results (M = -0.16)
  │ L2 Soft Results (M = -0.60)
  │
  │ L1 Hard Discussion (M = -1.43)
  │ L1 Soft Discussion (M = -1.44)
-1│ L2 Hard Discussion (M = -1.75)
  │ L1 Hard Introduction (M = -1.82)
  │
  │ L1 Soft Introduction (M = -2.18)
  │ L2 Soft Discussion (M = -2.23)
-2│ L2.Hard Introduction (M = -2.45)
  │ L2 Soft Introduction (M = -2.55)
  │
  │
```
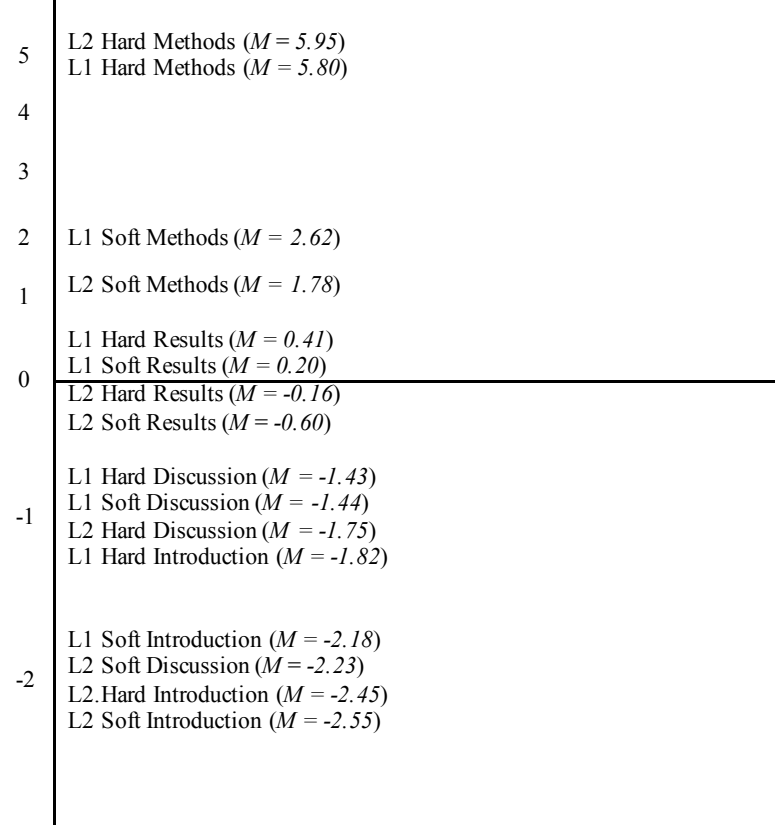
Figure 6.5. Dimension 3 mean scores for the IMRD sections written by L1 and LX authors across hard and soft domains (F = 81.8, p = .<0001, R² = .43)

PAST-TIME PROCEDURAL DISCOURSE

5		Hard Methods (M = 5.88)	
4	Methods (M = 4.02)		L1 Methods (M = 4.18)
3			L2 Methods (M = 3.83)
2		Soft Methods (M = 2.20)	
1			
0		Hard Results (M = 0.12)	L1 Results (M = 0.31)
	Results (M = -0.04)	Soft Results (M = -0.19)	L2 Results (M = -0.38)
-1	Discussion (M = -1.72)	Hard Discussion (M = -1.59) Soft Discussion (M = -1.84)	L1 Discussion (M = -1.44) L2 Discussion (M = -1.99)
-2	Introduction (M = -2.26)	Hard Introduction (M = -2.13) Soft Introduction (M = -2.36)	L1 Introduction (M = -2.00) L2 Introduction (M = -2.50)

Figure 6.6. Dimension 3 mean scores for different register categories

Finally, Dimension 3 shows no substantially difference between L1 and LX authors in their use of the features grouped on this dimension (see Appendix C, Table C8).

Altogether, the linguistic function underlying Dimension 3 appears to not only distinguish the discourse of Methods from those of other sections but also highlight disciplinary specificity in fulfilling the communicative purposes of this section.

6.3.5. Synthesis

In the previous sections, I provided descriptions of the three dimensions and showed that they are separate in terms of their linguistic characterizations and the functional construct they represent. I have also briefly considered the characterization of register categories with respect to each dimension. However, characterizing registers based on any of the three dimensions in isolation would potentially result in an incomplete description of their situational characteristics (see discussion in Biber, 1986; Biber, 1988, p. 164). Consider the description of IMRD sections with respect to the three dimensions, for example. Dimension 3 shows that Methods sections are characterized by their highly narrative, procedural discourse. However, looking back at the patterns of results revealed by Dimension 1, we find that the discourse of Methods sections is also marked for having a high informational focus. Thus, if the characterization of Methods sections was considered only in terms of one dimension, the description of their situational characteristics would be incomplete.

Figure 6.7 summarizes the results of our MD analysis for IMRD sections by plotting the mean score of each section against the overall corpus mean with respect to the three dimensions. The plotted patterns show that Introduction sections have scores above and below the average for Dimension 2 (conceptual vs concrete) and Dimension 3 (past-time procedural), respectively. Mean scores for Methods are substantially below the corpus mean for Dimension 1 (evaluative and elaborated vs informational), and above the average for

Dimension 3. Results sections have relatively intermediate scores for Dimension 1 and 3, but they are markedly below the average score along Dimension 2. And finally, Discussions are marked by their high mean scores on Dimension 1.
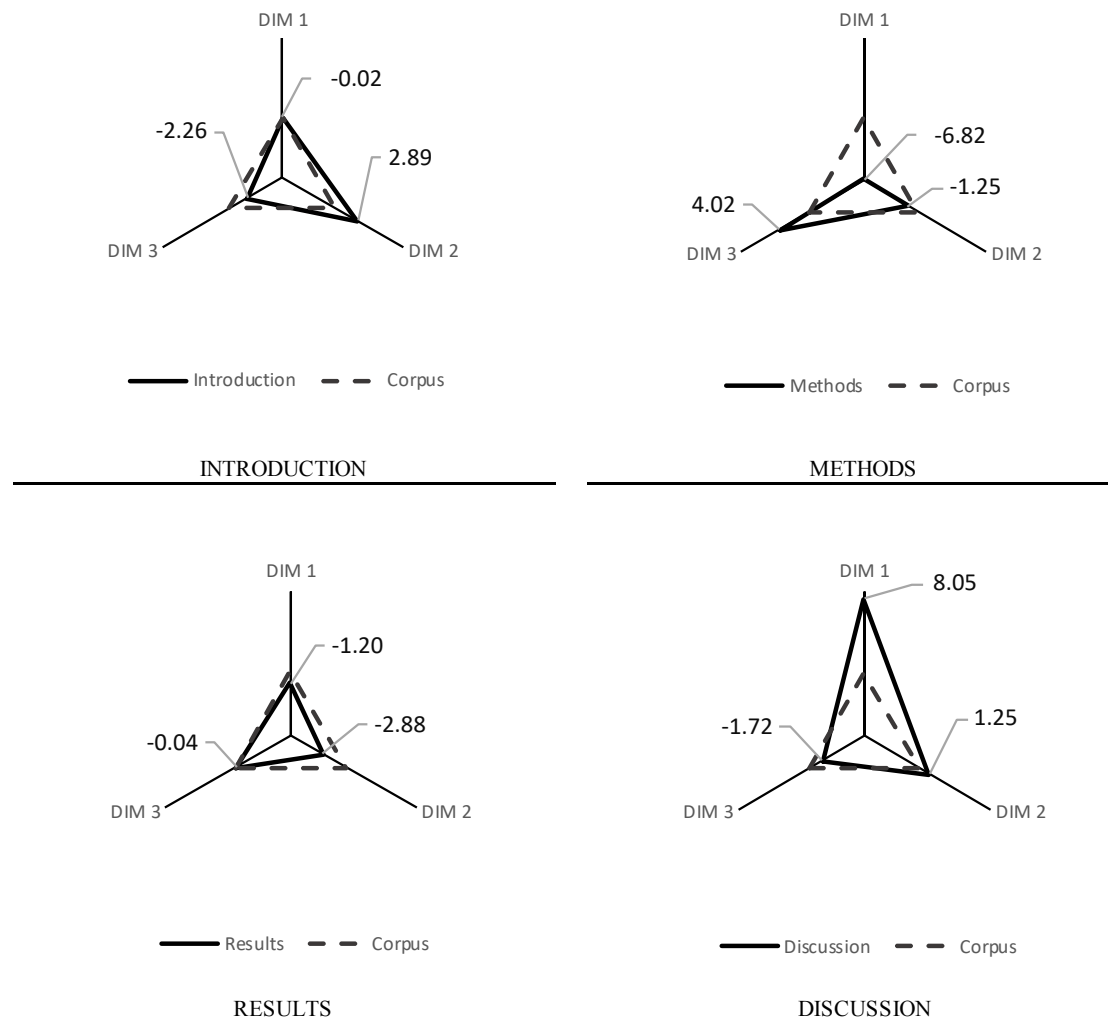
Figure 6.7. Plots of mean dimension scores for sections across the three dimensions

On the basis of these results, it is possible to provide a preliminary typology of research writing in IMRD sections with respect to the three dimensions and their corresponding linguistic features. This characterization is presented in Table 6.2.

Table 6.2. *A preliminary characterization of research writing in IMRD sections*

	Dimensions	Linguistic Features	Situational characteristics
Introduction	Dimension 2 ++ Dimension 3 --	abstract, cognitive, and human nouns; nominalization; 3rd person pronouns; communication verbs; attributive adjectives;	A conceptual, abstract, and expository discourse, characterized by precise choices of value-laden vocabulary
Methods	Dimension 1 --- Dimension 3 ++++	common nouns; nouns as pre-nominal modifiers; activity and aspectual verbs, time adverbials; past tense verbs; agentless passives	An informational, compressed, and procedural discourse, characterized by its focus on the narration of activities and past events while eliding the role of the researcher
Results	Dimension 2 --	concrete/tangible and quantity nouns, prepositions	A form of discourse that focuses on explanation of concrete and demonstrable evidence and rarely involves descriptions of abstract theoretical concepts
Discussion	Dimension 1 +++	predicative adjectives; general adverbs; demonstrative and *it* pronouns; verb BE; present tense verbs; necessity/obligation, possibility, prediction modals; that-clauses: controlled by stance adjectives and stance verbs; to-clauses: controlled by stance verbs; adverbial conjunctions	An evaluative and elaborated discourse, characterized by coherently patterns of convincing arguments and reasoning focused on the evaluation of findings and anchoring them in robust theoretical interpretations

++/--: ½ standard deviation above or below the mean; +++\---: ¾ standard deviation above or below the mean; ++++/----: 1 or more standard deviation above or below the mean

However, while such a functional characterization can provide useful insights into writing across sections of research articles, it does not account for the impact of disciplinarity on research writing in these sections. The results reported in the previous sections show that distribution patterns of IMRD sections are strongly influenced by disciplinary specificity across all three dimensions. Taking the mean dimension score of each section as a benchmark, Figure 6.8 allows for a comparison of these patterns with respect to the three dimensions.
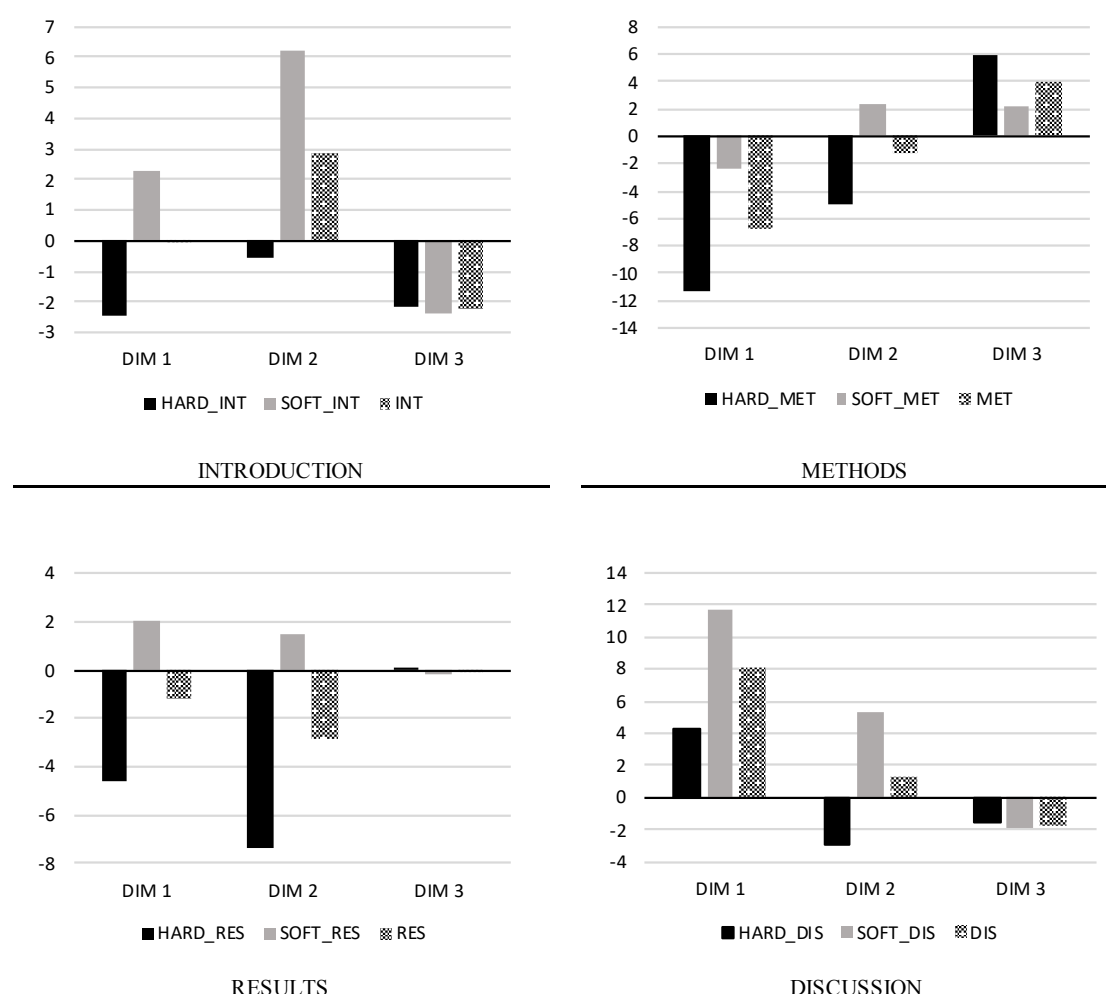
Figure 6.8. Plots of mean dimension scores for hard and soft fields across the three dimensions

Figure 6.8 shows that, on Dimension 1, Introductions in Humanities and Social sciences are characterized by their evaluative and elaborated discourse, while these sections in the hard fields typically have an informational focus. In addition, only the Introductions written in the soft fields are marked as being highly conceptual and abstract (Dimension 2), suggesting that the discourse characteristics of these sections can considerably vary depending on disciplinary preferences. For Methods, while the discourse of these sections in the sciences can be characterized as being informational (Dimension 1), event-driven/concrete (Dimension 2), and procedural (Dimension 3), such a characterization does not seem to be very well-suited for describing the prose style of Methods in the soft fields. Similar

disciplinary influences can be found by looking at the patterns shown for Results sections: these sections in Humanities and Social sciences have positive scores on both Dimensions 1 and 2, which contradicts the patterns shown for the sciences and the overall mean scores of Results sections along these two dimensions. This suggests that the characterization proposed in Table 6.2 may not prove accurate for describing the discourse of Results in the soft fields, and that writing these sections in Humanities and Social sciences, as opposed to the hard fields, may also rely on evaluation/elaboration and description of conceptual, abstract entities. Finally, Discussion sections in the hard disciplines have a considerably lower mean score on Dimension 1, compared to the overall mean score of these sections on this dimension. Moreover, the patterns shown for Discussions along Dimension 2 suggest that discussion of findings in the sciences often involves explanation of concrete evidence rather than interpretation of theoretical constructs (as is often the case in Discussions written in Humanities and Social sciences). This implies that discussion of research findings in sciences writing, in comparison to writing in the soft fields, is more empiricist and less interpretative in nature. Altogether, the patterns plotted in Figure 6.8 indicate that describing the situational characteristics of research writing across IMRD sections with no consideration given to possible disciplinary influences could result in tentative conclusions at best.

In addition to investigating the role of disciplinarity and intratextuality in research writing, the present study also explored L1- and LX- English scholars' writing practices across sections of research articles. The results of our MD analysis showed that only one dimension was significantly associated with systematic differences between L1 and LX expert writing. Specifically, discussions written by L1-English authors showed higher positive mean scores than those written by LX-English writers on Dimension 1. As noted earlier, the functional construct underlying the positive pole of this dimension can be described as being associated with discourse practices primarily concerned with evaluation

and elaboration. This finding suggests that there are considerable differences between L1 and LX expert writing in the Discussion section, especially with respect to fulfilling the communicative practices related to the evaluation and elaboration of research findings. Previous research has shown the Discussion section is characterized by rhetorical purposes that involve interpretation, elaboration, and validation of research findings, with the aim of convincing the reader to assent to a particular knowledge claim (e.g., Basturkmen, 2009; Swales & Feak, 2012; Yang & Allison, 2003). Lillis and Currey (2010) found that fulfilling such rhetorical functions is particularly challenging to non-Anglophone authors, and that these authors often find writing Discussion sections more difficult compared to other sections such as Methods. While I caution against interpreting the evidence presented here from a perspective of LX deficiency or difficulty, the results suggest that the nuances of writing Discussions and the communicative demands associated with evaluation and elaboration of research findings have the potential to shed light on important differences in research writing practices of L1-and LX-English scholars. It can be argued that such comparative data can put anecdotal evidence on L1-LX differences in expert writing into a more empirical perspective and validate assumptions underpinning such a dichotomy (also see Curry & Lillis, 2019; Hyland, 2016; Martinez, 2018).

6.4. Conclusion

Scholarly knowledge in contemporary academia is created and communicated primarily through writing for research publication. This particular form of knowledge dissemination takes place on a global scale in today's academia, involving more than eight million researchers around the world (see Ware & Mabe, 2015). It is now almost a truism to say that the vast majority of scholarly activities related to knowledge creation and its circulation in the modern academic system have close ties to research writing for publication purposes. Due

to its distinct linguistic characteristics and arcane discourse conventions, this form of writing has been known to pose great challenges to researchers, especially those at the beginning stages of their scholarly career. The present study set out to investigate the linguistic particularities and conventional discourse style of research writing, as reflected in the use of various lexical and grammatical features. For this purpose, a multi-feature/multi-dimensional approach was adopted to inductively identify co-occurring linguistic features of research writing and describe their underlying discourse functions. Our analysis revealed three fundamental linguistic dimensions which underlie the discourse practices of research writing across academic fields. These dimensions were tentatively labeled 'Evaluative and elaborated discourse vs. Informational focus', 'Conceptual and abstract vs. concrete language', and 'past-time procedural discourse'. The results further showed that there are substantial differences in the discourse style of research writing in sections of articles with respect to these dimensions. It was also found that these linguistic particularities can be strongly affected by disciplinary preferences, and that any characterization of research writing with no careful consideration given to disciplinarity would be incomplete. The results of the present study also showed that the writing practices of L1-and LX-English scholars when discussing research findings can be considerably different. Future research is required to further investigate these variations in L1 and LX expert writing. It can be argued that the functional parameters revealed by the first Dimension of variation identified in the present research can provide a useful starting point for a comparative study of such differences.

In sum, it appears that linguistics characteristics and communicative functions defining writing for research publication are complex and multi-faceted. Thus, comprehensive conclusions concerning the discourse characteristics of this genre of writing would need to be based on linguistic and functional descriptions from multiple perspectives.

Future research could adopt such a multi-perspective approach to shed further light on various aspects of research writing in academic contexts.

6.5. Acknowledgements

Chapter 7

Conclusion

7.1. Overview of the chapter

The central aim of this research has been to shed light on the arcane conventions of research writing by highlighting the ways in which meaning is constructed and conveyed through various linguistic means in academic fields. To accomplish this goal, three separate studies were conducted to examine the language of research writing from multiple perspectives, investigating the discourse characteristics of this writing genre at three linguistic levels: lexis, multi-word expressions, and general lexico-grammatical items. The first two studies were designed to investigate the role of disciplinarity and intratextuality as key parameters of variation in the use of single and multi-word items in research writing. The purpose of the third study was to not only account for the influences of these key factors on the use of various lexical and grammatical features of research writing, but to also examine possible variation arising from differences in writing practices of L1 and LX-English scholars.

The nature of the research goals of these three studies required representative corpus data that would allow for such linguistic investigations. In addition, investigating the language of research writing from multiple perspectives necessitated the use of various research methods and data analysis techniques. The goal of the present chapter is to summarize and further discuss the findings that emerged from the corpus-based analytical methods used in this research. In doing so, the chapter first highlights the advantages of the multi-perspective research approach adopted in this study and discusses the benefits of such methodological triangulation for corpus-based research in general. It then provides a summary of main research findings and further discusses convergent evidence arising from the three studies

reported in previous chapters. The chapter concludes with a discussion of the implications of this research for understanding the distinct linguistic characteristics and the discourse conventions of writing for research publication across academic fields.

7.2. Representing the language of research writing through corpus data

Corpus-based language studies that attempt to examine the linguistic characteristics of a particular form of language use should ideally follow three important methodological steps (see Baker & Egbert, 2016; Biber & Conrad, 2009; Biber & Egbert, 2016):

> i. perform observational assessments of the target domain to identify the situational characteristics of the target texts that are to be included in the corpus
>
> ii. conduct principled sampling procedures to compile a representative sample of texts from the target domain
>
> iii. use various corpus analysis techniques to not only identify but also validate, through cross-checking, the identified patterns of language use

The above methodological procedures can increase the generalizability of research findings, allowing the researcher to anchor the observed linguistic patterns in more robust theoretical interpretations. Although all these methodological steps seem to be important in arriving at valid and verifiable linguistic results, most corpus-based studies often focus only on the second step, rarely acknowledging or employing the first and third steps (see Baker & Egbert, 2016; Biber & Egbert, 2016). In conducting the research reported in this dissertation, however, I strived to give careful considerations to all the above methodological procedures. Specifically, in order to systematically implement the first step, surveys of research articles published across academic fields were conducted to identify the important parameters that could be considered as situational characteristics of research writing. In doing so, a list of well-established disciplines was first developed on the basis of their epistemological

orientations. An inductive survey of high-ranking journals was then conducted to determine the type of articles commonly published in each disciplinary field: full-length articles in various volumes of each journal were carefully surveyed and only those that exhibited distinct characteristics of research writing were selected (i.e., empirical research articles). Empirical papers in each discipline were then examined to determine their internal structures. These articles were also classified based on the method for authorship categorization described in Chapter 3. Undertaking the first methodological procedure helped ensure the representativeness of the corpus from two perspectives: external (situational) and internal (linguistic) representativeness (see Biber, 1993; McEnery, Xiao & Tono, 2006, p. 14). External representativeness relates to the extent to which a sample of texts is extracted from the range of text types representing the target discourse domain (Biber, 1993). By surveying the range of article types commonly published across academic fields and focusing on empirical research articles, as the article type most representative of research writing for publication (see Hyland, 2016), the corpus was safeguarded against a possible violation of external representativeness. Internal representativeness is viewed as the extent to which the text samples in a corpus account for the range of internal linguistic variation within the target discourse domain (Biber, 1993). This view is based on the premise that the linguistic characteristics of language use within a given register are not always homogeneous and can vary due to internal inconsistency in the communicative purposes of the discourse under investigation (see Biber & Conrad, 2009). As was mentioned in Chapter 3 (Section, 3.2.2), prior to sampling, articles were carefully examined to determine their textual layouts and internal organization. This ensured that all the articles included in the corpus contained a similar textual organization (i.e., the content of the standard IMRD sections) and that they could represent possible linguistic inconsistencies within empirical research articles.

Following the implementation of the first step, the extraction of articles was carried out based on a stratified random sampling procedure through which five articles were sampled from different issues of a journal in each year. Through this sampling procedure a roughly equal number of articles from different journals in each discipline were selected so as to avoid possible journal influences on writing style. In addition, care was also taken to reach comparable word counts for the selected disciplines. Each extracted article was then thoroughly examined to ensure that it conformed the operational criteria used in implementing the first step. Undertaking the first two methodological procedures resulted in a principled, balanced corpus, which was then analyzed through various corpus-based analytical approaches, systematically implementing the third methodological consideration described above. The following section discusses these analytical techniques and highlights the key advantages of such a multi-method approach for identifying and verifying linguistic findings.

7.3. Research approaches, their advantages and main findings

The research approach used for conducting the first two studies of this dissertation was inductive and bottom-up. Through this research approach, various linguistic patterns related to the use of single and multi-word items emerged from the corpus with little theoretical assumptions guiding their identification. The third study, however, used an approach which can be considered as having characteristics of both deductive and inductive approaches. On the one hand, the approach adopted in this study can be considered deductive because it relies on a set of lexical and grammatical features annotated by an automatic grammatical tagger (developed and revised by Biber). On the other hand, it can also be seen as an inductive approach since it uses the statistical technique of factor analysis which analyzes the statistical co-occurrence patterns of linguistic items from the bottom up, determining the grouping of

these features based on their systematic co-occurrence in discourse. In each of these studies, texts in the corpus were explored from different perspectives, using various methods of data analysis. The following sections summarize and briefly discuss the analytical methods and main findings related to each of the three studies conducted in this dissertation research.

7.3.1. Axes of variation in the use of single-word vocabulary

In Study 1, Gries's (2008) deviation of proportions statistic was used to assess the extent and nature of variability in vocabulary use in writing for research publication. The analysis showed that the majority of the high-frequency vocabulary used in research writing is not widely shared among academic disciplines, indicating that the language of this genre of writing is far from being homogeneous and is characterized by highly specialized and field specific discourses. It was also found that research writing in the sciences demands the knowledge of a type of vocabulary that has a narrow range of use and applicability. This analytical technique was also used to examine possible internal variation in terms of vocabulary use across main sections of research articles. A considerable degree of unevenness in vocabulary use was found across the main sections of research articles, providing evidence in support of the above argument concerning the importance of considering possible internal inconsistency within linguistic registers. Further patterns of use were identified by looking at these results with respect to disciplinary influences: a) science writing was found to heavily rely on narrowly dispersed vocabulary across all sections, verifying the patterns of vocabulary use by science writers identified earlier and b) explaining research methods in the sciences was found to require a vast repertoire of a type of vocabulary which can have a very narrow range of use in research writing in general.

In addition to Gries's (2008) dispersion statistic, Study 1 also used the corpus-based method of keyword analysis in conjunction with the statistical technique of hierarchical

agglomerative clustering. This analysis provided a systematic mapping of the commonality and variations between disciplines, indicating that the degree of homogeneity in the use of high frequency vocabulary that plays a key role in knowledge creation in academic fields not only varies between hard and soft disciplines but also within these two broad domains. While the confluence of certain disciplines in terms of key vocabulary was found to be stronger than others, certain disciplines (e.g., applied linguistics) were found to show a very low degree of homogeneity with other fields, providing an indication of their outlier status in the use of high-frequency keywords. These results were further verified by investigating possible commonalties between the sections of research articles: the degree of overlap between sections was also found to be governed by disciplinary conventions. Specifically, it was shown that while the discussion of results in sciences writing can be primarily based on reiterating their experimental basis and anchoring them in empirical observations, justification of findings in Social Science and Humanities is often realized through the lens provided by a theoretical basis established in the Introduction section. Overall, these results illustrate how standards and conventions of scholarly activities in academic fields can govern the delineation of authors' linguistic decisions at the most basic linguistic levels (e.g., the lexis). Moreover, the various methods adopted in this study represent a handful of ways that multi-method approaches can be systematically used in linguistic research, highlighting the benefits of triangulating results obtained from more than one method of data analysis. The above findings also serve to illustrate that investigating patterns of language use at different levels of granularity can result in more generalizable conclusions.

7.3.2. Variation in form-function connection of multi-word expressions

Study 2 reported in this dissertation also relied on a mixed-methods design to explore variation in the form-function mappings of multi-word expressions in research writing. To

examine such differences, a two-stage mixed-methods research design was employed, through which linguistic patterns were first quantitatively analyzed and then further examined by a qualitative appraisal. Similar to Study 1, the identification of linguistic patterns was carried with minimal a priori assumptions. Specifically, recurring multi-word items in the corpus were identified and their occurrence rates were calculated using computer programs developed for the particular purposes of the study. The identified sequences were then classified according to their structural correlates and the communicative functions they commonly serve in discourse. The results from the structural analysis showed systematic patterns of disciplinary variation in the use of noun phrases, with sciences research writing heavily relying on using different types of sequences with noun phrase structures. In contrast, it was found that research writing in Social Sciences and Humanities fields tend to rely on the use of prepositional and comparative expressions that work to situate arguments within a particular context (e.g., *on the basis of*) and establish links between elements of discussion (e.g., *as well as the*). Writing Methods in the soft fields was also marked by frequent use of verbal structures that incorporate clausal fragments, reflecting a more discursively elaborate presentation of research methods in these fields than in the sciences. The results from the functional analysis corroborated and verified the patterns revealed by the structural analysis, indicating a strong relationship between the structural form of the identified expressions and their discourse functions in research writing. For instance, it was found that, in writing their methodologies, writers in Social Sciences and Humanities tend to make extensive use of text- and participant-oriented expressions, resulting in a discourse style which is more elaborated/interpretative and less empiricist in nature. These expressions were found to be mostly verb-based and prepositional phrases, used to construct meaning related to elaboration and evaluation.

The results of the qualitative analysis showed that, in writing Introductions and Discussions, writer in the sciences tend to make extensive use of distinctive research-oriented expressions that are used to describe and specify physical and tangible aspects of research objects, suggesting that writing these sections in these fields is more focused on explanation of concrete and demonstrable evidence rather than explicating their theoretical basis. In contrast, it was found that writing Introductions in the soft fields was marked by a heavy reliance on intangible framing expressions, employed to specify abstract and intangible characteristics of conceptual entities. The qualitative analysis also revealed key differences in the ways in which the communicative aims of Methods and Results are realized through certain distinctive expressions across disciplines. It was shown that writers in Social Science and Humanities tend to make frequent use of attitudinal/modality expressions to provide tentative support for their choice of a particular research method, present their results in a less assured manner, and avoid overgeneralizations of their findings. Such expressions allow authors to ensure that their arguments, views and evaluations about the methodology used or the findings obtained in their research are expressed with an appropriate level of confidence and commitment. The patterns of use uncovered by the qualitative analysis suggest that adopting such discursive strategies when writing Methods and Results sections is more common in the soft fields than in the sciences. Taken together, the findings from the second study demonstrated that there is a strong form-function connection between multi-word items and the communicative purposes they fulfill in research writing, and that this connection is strongly affected by disciplinary conventions across main sections of research articles. In addition, the specific methods adopted in this study show the methodological strength of mixed-methods research designs in providing cross-checked, triangulated linguistic data with an acceptable level of validity and reliability.

7.3.3. Dimensions of variation in the use of lexico-grammatical features

In conducting Study 3 of this dissertation, Biber's (1988) MD analysis was used. As was mentioned in Chapter 2, this analytical technique can be employed to analyze a wide range of lexical and grammatical features of a particular domain of language use. MD analysis can be used through two different approaches: additive MD approach and full MD approach. For Study 3, the latter approach was adopted, since a full MD analysis has the potential to lead the researcher to hitherto unknown patterns of linguistic variation (see Chapter 2, Section 2.3.4.3 for further discussion). In addition, the full MD approach is appropriate when investigating such a specialized discourse domain as research writing that is characterized by a highly conventional and specific prose style. Thus, a full MD analysis has the potential to shed light on the distinctive characteristics of research writing through the identification of the linguistic patterns specific to this particular domain of language use. In this approach, the notion of linguistic co-occurrence is given a formal status through a multivariate statistical technique (i.e., factor analysis/principled component analysis), which quantitatively identifies the statistical co-occurrence patterns of lexico-grammatical features in a corpus (see Chapter 6, Section 6.2 for further discussion).

The full MD analysis is based on a mixed-methods research design that is cyclical rather than unidirectional. Specifically, the analysis follows a cyclic process whereby a qualitative appraisal of quantitatively defined patterns is conducted to inform quantitative investigations of the data, which in turn lead the researcher to detailed qualitative explanations and linguistic theories about the discourse domain under investigation (the discussion throughout Section 6.3 of Chapter 6 is an example of this process). Conducting the first part of this process resulted in the identification of three specialized functional constructs defining the discourse of research writing: 'Evaluative and elaborated discourse vs. Informational focus', 'Conceptual and abstract vs. concrete language', and 'past-time

procedural discourse'. The interpretation of these functional dimensions was carried out on the basis of the statistically identified co-occurrence patterns of lexico-grammatical features. Following this, similarities and differences among various register categories of the corpus were quantitatively assessed with respect to each of these constructs. The results from this part of the process were then used as the basis for more detailed qualitative explanations of the observed patterns, which was conducted with respect to all the three identified dimensions of variation. Through this process, a preliminary typology of research writing and its linguistic characteristics across IMRD sections was proposed. However, it was shown that such a characterization without considering possible disciplinary influences was incomplete and tentative at best, indicating that knowledge creation through the use of various lexico-grammatical features is governed by disciplinary preferences, and that such influences are present in overall patterns of language use. It was also found that the writing practices of L1- and LX English scholars were substantially different when fulfilling communicative purposes related to evaluation and elaboration of research findings. This particular difference was observed in the Discussion section, which is characterized by conventional discourse practices such as interpretation of results and providing convincing arguments through evaluation and elaboration of findings.

In sum, the analytical approaches used to conduct Study 3 of this dissertation provide an indication of how different methods of data analysis can inform one another through a cyclic process. The findings that emerged from such a research process highlighted various linguistic aspects and nuances of writing for research publication. The picture that emerged from these findings points to the conclusion that textual characteristics of research writing are multidimensional, hence requiring a comprehensive methodology that could capture various aspects of such a complex form of language use.

7.4. Further discussion of findings

In the modern academic system, research writing for publication is arguably one of the primary means of disseminating knowledge, involving more than eight million researchers around the world (see Ware & Mabe, 2015). It is through this particular form of knowledge production that a rich body of scientific knowledge about a given phenomenon is accumulated and transmitted to its target audience. The research reported in this dissertation has attempted to provide a comprehensive view of the language of this specific sub-register of academic writing. In doing so, a multi-perspective approach was adopted through which patterns of language use in this genre were investigated from three different linguistic perspectives: lexis, multi-word expressions, and general lexico-grammatical items. Each linguistic perspective captured a unique aspect of the particularities distinguishing the language of research writing across disciplines. The picture of research writing that has been painted by this study shows that authors' linguistic choices at various levels of language use, such as lexis, multi-word expressions, and more complex lexico-grammatical linguistic structures, have close ties to the epistemological orientations of their disciplinary community and the field-specific communicative functions they aim to fulfill at various stages of composing their research papers. This suggests that discourses produced as a result of such linguistic choices are not monolithic, homogeneous, or uniform. This recognition raises fundamental questions about the idea of generality informing most EAP studies and the field of EAP in general. While EAP has taken different forms and directions since its inception in the 1960s, the founding premise of the field is that particular areas of academic discourse are relatively homogeneous in nature, such that learning the linguistic features representing this homogeneity would help students gain control of the conventions of academic communication and successfully navigate the challenges involved in acquiring such a competence (Blue, 1988; Hutchinson & Waters, 1987, p. 165-166; also see Jordan, 1997).

The present research has consistently shown that such a theoretical stance can hardly be considered a definitive doctrine for EAP. More specifically, the evidence presented in this study has shown that the linguistic characteristics of one of the most central genres of academic communication (i.e., research writing) are so diverse and disparate that providing an overarching, unified characterization of its discourse for students is virtually impossible. This specificity in language use is inherently linked to the diversity that exists among academic communities, their intellectual practices, and the different forms of knowledge they produce (see Chapter 2).

The present study has shown that the communication of scholarly knowledge through research writing is a highly conventionalized genre in which effective transmission of meaning is dependent on familiarity with the linguistic expectations of the target readership. Gaining control of the unique discourse style of this genre of writing demands a strong understanding of linguistic particularities that characterize research writing and differentiate it from other written genres. Developing such a specialized competence typically requires engaging in a protracted process that involves prolonged exposure to the linguistic conventions governing the distinctive prose style of research writing (see e.g., Biber et al, 1999; Curry & Lillis, 2014; Lillis & Curry 2010). The majority of such linguistic conventions, however, are strongly influenced by the diverse and heterogeneous nature of knowledge and its production across academic fields. Acquiring the literacy skills associated with research writing appears to be symbiotically linked with the ability to recognize the boundaries of the conventions defining the discourse practices of a given academic community. Thus, gaining control of research discourses involves facing the dual task of mastering a particular form of knowledge communication (i.e., research writing) while also gaining familiarity with the linguistic demands associated with field-specific conventions.

A major reason for exploring the language of research writing, then, is that it allows us to unpack the textual and disciplinary norms governing the act of writing for research publication across academic fields (see Hyland 2009). Such conventions carry assumptions about the ways in which epistemological orientations of disciplinary communities, and their routine scholarly practices, imbricate the construction of knowledge in contemporary academia. From a linguistics perspective, this means that the discourse conventions of disciplinary communities have the potential to shape, affect, and/or change the general norms of research writing. For example, the research reported in this dissertation showed that, although the extensive use of nouns, nominal phrases, and embedded phrasal structures is a unique characteristic of research writing in general (see Biber & Gray, 2016), extensive use of verbal structures that incorporated clausal fragments was a distinctive characteristic of research papers written in Humanities and Social sciences, reflecting a discursively elaborate presentation of research in these fields. Evidence of this sort suggests that linguistic properties of research writing are best described in relation to the disciplinary fields in which they are commonly used and that characterization of research discourses in absolute terms could potentially result in tentative conclusions at best (for specific discussions of similar findings, see the discussion sections for 2 and 3; Sections 5.4 and 6.3, respectively).

The research reported in this dissertation has also shown that the discourses created in the process of research writing are multi-faceted and complex constructs, often strongly influenced by the combined effects of both social and contextual factors. More specifically, the findings from the three studies have consistently shown that research writing is profoundly affected by the interaction between disciplinary specificity and internal shifts in purpose within texts (i.e., intra-textual variation). This implies that a characterization of research writing across disciplinary fields without careful consideration given to the specific ways in which they fulfill the communicative functions associated with text-internal shifts

within research articles would be incomplete (see the discussion of findings for Study 1, 2 and 3; Sections 4.4, 5.4, and 6.3, respectively). Study 3 of this dissertation also showed that similar situations could arise if we attempt to provide a characterization of intra-textual variation within research articles with no consideration given to the impact of disciplinary specificity (see discussion in Section 6.3.5 of Chapter 6). This strong connection between disciplinarity and intratextuality indicates that the integration of such social and contextual parameters is a fundamentally important consideration for exploring the language of research writing (see Candlin & Hyland, 1999). Included in this view is a more sophisticated understanding that recognizes research writing as a situated activity that cannot be simply described as a discrete act of a writer (see Prior, 1998). From this view, the texts created in the process of research writing are complex constructs, representing not only their author(s) but also the social and contextual situations within which they occur. This in turn implies that comprehensive conclusions concerning the discourse characteristics of research writing would need to be based on linguistic and functional descriptions from multiple perspectives.

In addition to investigating the role of disciplinarity and intratextuality in research writing, the present research also explored possible variation in the writing practices of L1- and LX- English scholars. The results showed that discussions written by L1-English authors were considerably different from those of their LX-English peers in terms of fulfilling the communicative practices related to the evaluation and elaboration of research findings. This suggests that perhaps communicative situations that require evaluations and elaborations could be viewed as potentially important sites for the study of linguistic differences between the writing practices of L1- and LX- English scholars. Research on exploring writing for research publication in a second or foreign language has drawn on various approaches, often using qualitative research methods such as surveys and focus groups (see Lillis & Curry, 2016). While anecdotal evidence about L1-LX differences in research writing abound, little

comparative linguistic data have been offered in evidence. Conclusions drawn from corpus-based analyses have the potential to validate assumptions underpinning L1-LX differences in research writing by putting anecdotal evidence on such a dichotomy into a more empirical perspective. The findings from this research show that the linguistic choices made by authors to perform the communicative purposes requiring elaboration and evaluation can provide a useful starting point for a comparative study of such differences. The findings also suggest that such a comparative study could provide important insights into L1-LX expert writing by exploring possible linguistic variation in the construction of argumentation in the Discussion section (also see Lillis & Curry, 2010). The discourses created in the process of writing Discussions often represent the use of language at its most delicate and eloquent level, reflecting the author's attempt at presenting justifications and evaluations as convincing and coherently organized arguments (see e.g., Basturkmen, 2009; Swales & Feak, 2012; Yang & Allison, 2003; for further discussion see Chapter 6). Such carefully crafted discourses seek to establish the credibility of knowledge claims, convince readers, or turn their attention to the necessity of a proposition through the logical force of the argument, representing the delicate ways in which research outcomes are effectively conceptualized as scientific knowledge through interpretations of empirical evidence. The results from the present research indicate that the nuances of writing Discussions and the communicative demands associated with this section of research article have the potential to shed light on important differences in research writing practices of L1-and LX-English scholars. These findings also imply that investigating research writing in a second/foreign language across specific sections of research article could provide a more nuanced view of the writing practices of EAL academics.

Altogether, the findings reported in this dissertation show how different approaches to the analysis of research articles and different perspectives on writing for research publication can be integrated to provide a better understanding of this preeminent genre of knowledge

communication in contemporary academia. In applied linguistics research, orientations

towards the study of research writing often considerably vary at various levels, such as the

methodological approaches used (e.g., the unite of analysis under study) or the empirical

object under investigation (e.g., whether the focus is on textual aspects, social and contextual

factors or variations related to writing practices of authors) (Lillis & Curry, 2016). The

present research, however, has attempted to make explicit the valuable links that can be made

between these orientations and show how different approaches and perspectives can be

integrated to offer a more comprehensive view of research writing.

7.5. Limitations of the research

In this research, careful consideration was given to situational characteristics of language use

in research writing. As was explained earlier (see Section 7.2), this involved performing

detailed observational assessments of research articles to identify the situational factors that

could potentially affect language use in the target domain (i.e., research writing). Conducting

such assessments was an important step in the methodology used in this research, because it

provided the opportunity to control for confounding and extraneous variables (see Chapter 3).

The information gathered from such assessments laid the groundwork for creating a set of

operational criteria for determining which articles should be included in the DCRA. These

criteria provided operational definitions for determining the type of research articles (e.g.,

empirical research papers), their organizational structure (e.g., the IMRD framework), and

the language status of their authors (see Chapter 3, Sections 3.2.1, 3.2.2 and 3.2.3). While

such criteria helped ensure the validity of the linguistic comparisons made in the research,

they also limited the scope of the study.

More specifically, the present research focused primarily on providing descriptions of

language use in empirical research articles. And so, theoretical articles were not included in

the corpus. This decision was made based on the observation that the theoretical articles were not as common as empirical research articles across academic fields. As discussed in Section 3.2.1 (see Table 3.1), research articles that included empirical data were found to be common in almost all disciplines. Hence, it was decided to focus primarily on this particular article type and linguistic characteristics of language use in theoretical articles were not investigated in the present research. Therefore, it is important to note that the findings reported in this dissertation should not be generalized beyond the target domain (i.e., empirical research articles) investigated in this research.

Another limitation is related to the focus on IMRD sections as the target structural organization. As was mentioned in Chapter 3, articles that included a unique and distinctive sectioning format were excluded. The rationale for this decision was that most empirical research articles across different academic fields contained the content of a standard four-part organization (Introduction, Methods, Results, Discussion; IMRD). And so, for comparison and consistency purposes, it was decided to focus on this particular sectioning format. However, as was shown in Chapter 3 (Section 3.2.2), some research papers (e.g., mathematics papers) may not follow such a four-part structure. Future research is required to shed light on the linguistics characteristics of research writing in such articles.

A final limitation pertains to the focus on fully separated Results and Discussion sections. As was explained in Section 3.2.2 of Chapter 3, the rationale for this decision was that R&D sections typically fulfill distinct communicative purposes and perform different discourse functions in research articles. However, research articles (especially in Chemistry) may also include integrated R&D sections (see Section 3.2.2). But these articles were excluded from further analysis in this research to maintain the distinction between these two sections and acknowledge the significance of their distinct communicative purposes and

discourse functions. Future research is needed to investigate the linguistic characteristics of research writing in fully integrated Results and Discussion sections.

7.6. Implications and future research

The conventions of scientific writing in academic fields can present daunting challenges for students and those researchers who are at the beginning of their scholarly career. The view of research writing that has emerged from the present research has implications for materials developers and ESP instructors. Understanding variations in the language of research writing at various linguistic levels can help materials developers provide instructions and effective teaching materials that focus on raising students' awareness of communicative conventions in their academic fields. This knowledge could also benefit ESP instructors by helping them design their courses with teaching materials particularly relevant to students' content areas, providing them with a feasible starting point for understanding the linguistic norms of their discourse community.

The present dissertation study has important implications for future corpus-based research. As was noted earlier in this chapter, a secondary goal of this dissertation was to highlight the methodological benefits of multi-method and multi-perspective approaches in linguistic research. For this purpose, data-driven patterns of language use at three linguistic levels were identified, triangulated and verified through various corpus-based analytical techniques so as to provide multiple, complementary perspectives on the discourse of research writing across disciplines. These data analyses were based on a balanced, principled corpus of empirical research articles from a wide range of disciplines: biology, chemistry, dentistry, physics, mechanical engineering, applied linguistics, business, management, politics, and sociology. The analyses relied on quantitative and qualitative assessments to offer a better understanding of the identified patterns. The findings from this research have

shown that gaining multiple insights and perspectives is important for defining such a multi-layered form of language use as research writing. The key strength of such multi-perspective approach is that it allows the researcher to triangulate the emerged patterns using different techniques of data analysis. Such a methodological triangulation can yield results that can lay greater claim to generalizability, helping the researcher to anchor the emerged observations in more robust linguistic interpretations. Future research could adopt such methodological approaches to shed further light on various aspects of research writing. It is also possible to expand the multi-perspective approach used in this study by including other linguistic features such as grammatical complexity features. Further research of this kind appears to hold great promise for enhancing our understanding of the language of one of the most important forms of knowledge communication in contemporary academia (i.e., writing for research publication).

References

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes*, *31*(2), 81-92.

Afros, E., & Schryer, C. F. (2009). Promotional (meta)discourse in research articles in language and literary studies. *English for Specific Purposes, 28*, 58-68.

Aktas, R. N. & Cortes, V. (2008). Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes 7*. 3–14. DOI: 10.1016/j.jeap.2008.02.002

Anderson, A., & Valente, J. (2002). *Disciplinarity at the Fin de Siècle*. Princeton University Press.

Baker, P., & Egbert, J. (2016). *Triangulating methodological approaches in corpus linguistic research*. Routledge.

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh 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. DOI: 10.1016/j.jeap.2011.10.004

Bauer, L., & Nation, P. (1993). Word families. *International journal of Lexicography*, *6*(4), 253-279.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. University of Wisconsin Press.

Bazerman, C. (1994). *Constructing experience*. Carbondale IL: Southern Illinois University Press.

Becher, T. (1981). Towards a definition of disciplinary cultures. *Studies in Higher education*, *6*(2), 109-122.

Becher, T. (1989). *Academic tribes and territories: intellectual enquiry and the culture of the disciplines*. The Society for Research into Higher Education and Open University Press.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher education*, *19*(2), 151-161.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories*. The Society for Research into Higher Education and Open University Press.

Beck, I. L., McKeown, M. G., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples*. Guilford Press.

Bender, T. (1997). Politics, intellect, and the American university, 1945-1995. *Daedalus*, *126*(1), 1-38.

Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational Researcher*, *31*(8), 18-20. DOI: 10.3102/0013189X031008018.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman.

Biber, D (1993). Representativeness in corpus design. *Literary and Linguistic Computing 8*(4): 243–257. DOI: 10.1093/llc/8.4.243

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language 64*(2), 384-414.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (1994). An analytical framework for register studies. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 31-56). New York: Oxford.

Biber, D. (2006a). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D. (2006b). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, *5*(2), 97-116.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International journal of corpus linguistics, 14*(3), 275-311.

Biber, D. & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Biber, D. & J. K. Jones. (2009): "Quantitative methods in corpus linguistics", in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics: An International Handbook Vol. 2*. Berlin, New York, 1286-1304.

Biber, D., & Egbert, J. (2016). Using grammatical features for automatic register identification in an unrestricted corpus of documents from the open web. *Journal of Research Design and Statistics in Linguistics and Communication Science*, *2*(1), 3-36.

Biber, D., & Finegan, E. (2001). Intra-textual variation within medical research articles. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional Studies* (pp. 108-137). London: Longman.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2-20.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam/Philadelphia: John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at: lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(5), 371–405.

Biber, D., Egbert, J., & Keller, D. (2020). Reconceptualizing register in a continuous situational space. *Corpus Linguistics and Linguistic Theory*, *1*, 2-36.

Biber, D., Egbert, J., Gray, B., Oppliger, R., & Szmrecsanyi, B. (2016). Variation versus textlinguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In M. Kytö & P. Pahta (Eds.), *Handbook of English historical linguistics* (pp. 351-375). Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.

Biber, D., Reppen, R., Schnur, E., & Ghanem, R. (2016). On the (non) utility of Juilland's D to measure lexical dispersion in large corpora. *International Journal of Corpus Linguistics*, *21*(4), 439-464.

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psycholinguistics, 57*(3). 195–203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, *57*(3), 204-213. DOI: 10.1037/h0034699.

Blue, G. (1988). Individualising academic writing tuition. In P. Robinson (Ed.), *Academic Writing: Process and Product* (pp. 95–99). Modern English Publications & The British Council.

Braine, G. (1995). Writing in the natural sciences and engineering. In D. Belcher, & G. Braine (Eds.), *Academic writing in a second language: essays on research and pedagogy* (pp. 113–134). New Jersey: Ablex.

Brieger, N., & Pohl, A. (2002). *Technical English: vocabulary and grammar*. Heinle Cengage Learning.

Brown, R. (2004). Self-composed: Rhetoric in Psychology personal statements. *Written Communication*, *21*(3), 242–260. DOI: 10.1177/0741088304264338

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. DOI: 10.1016/j.jeap.2007.12.001

Bucher, R. & Strauss, A. L. (1961). Professions in process. *American Journal of Sociology,* 66, 325-34.

Charles, M. (2003). A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, *2*, 313–326.

Charles, M. (2006). The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics, 27*, 492-518.

Charles, M. (2007). Argument or evidence? Disciplinary variation in the use of the Noun that pattern in stance construction. *English for Specific Purposes, 26*, 203-218.

Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL Word families in medical research articles (RAs). *English for Specific Purposes, 26*, 502–514.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, *37*(6), 849-880.

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, *32*(2), 251-263.

Clark, B. R. (1983). *The higher education system*. Berkeley, CA: University of California

    Press.

Clark, B. R. (1987). *The Academic Life. Small Worlds, Different Worlds. A Carnegie*

    Foundation Special Report. Princeton University Press.

Conrad, S. (1996a). *Academic Discourse in Two Disciplines: Professional Writing and*

    Student Development in Biology and History. Doctoral dissertation, Northern Arizona

    University.

Conrad, S. (1996b). Investigating academic texts with corpus-based techniques: An example

    from biology. *Linguistics and Education, 8*, 299-326.

Conrad, S., & Biber, D. (2001). *Variation in English: Multi-dimensional studies*. Longman.

Corson, D. (1997). The learning and use of academic English words. Language Learning,

    47(4), 671-718.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples

    from history and biology. *English for Specific Purposes 23*(4). 397–423. DOI:

    10.1016/j.esp.2003.12.001

Cortes, V. (2013). The purpose of this study is to: connecting lexical bundles and moves in

    research article introductions. *Journal of English for Academic Purposes 12*(1). 33-43

    DOI: 10.1016/j.jeap.2012.11.002

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum. In V. Cortes

    & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor*

    of Doug Biber (pp. 197-236). Amsterdam: John Benjamins.

Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The encyclopedia of language and*

    linguistics (pp. 3168–3171). Oxford, UK: Oxford University Press.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Coxhead, A. (2019). Academic vocabulary. In S. Webb (Ed), *The Routledge handbook of vocabulary studies* (pp. 97-110). Routledge.

Coxhead, A., & Nation, I. S. P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge University Press.

Coxhead, A., Stevens, L., & Tinkle, J. (2010). Why might secondary science textbooks be difficult to read?. *New Zealand Studies in Applied Linguistics*, *16*(2), 37.

Crawford, W., & Csomay, E. (2015). *Doing corpus linguistics*. New York, NY: Taylor and Francis Inc.

Cruttenden, A. (1981). Item-learning and system-learning. *Journal of Psycholinguistic Research 10*, 79–88.

Curry M. J., & Lillis, T. (2017). Problematizing English as the privileged language of global academic publishing. In M. J. Curry & T. Lillis (Eds), *Global Academic Publishing: Policies, Perspectives and Pedagogies* (pp. 1-20). Multilingual Matters.

Curry, M. J., & Lillis, T. (2019). Unpacking the lore on multilingual scholars publishing in English: A discussion paper. *Publications*, *7*(2), 1-14.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics*, *14*(2), 159-190.

Dewaele, J. M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker'. *Applied Linguistics, 39*(2), 236-240.

Donald, J. G. (1990). University professors' views of knowledge and validation processes. *Journal of Educational Psychology*, *82*(2), 242-249. DOI: 10.1037/0022-0663.82.2.242.

Dudley-Evans, T. (1994). Variations in the discourse patterns favoured by different

    disciplines and their pedagogic applications. In J. Flowerdew (Ed.), *Academic*

    listening: Research perspectives (pp. 46-58). Cambridge: CUP.

Durrant, P. (2008). *High-frequency collocations and second language learning*. (Unpublished

    doctoral dissertation). The University of Nottingham, England.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for

    academic purposes. *English for Specific Purposes*, *28*(3), 157-169.

Durrant, P. (2014). Discipline and level specificity in university students' written

    vocabulary. *Applied linguistics*, *35*(3), 328-356.

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university

    student writing?. *English for Specific Purposes*, *43*, 49-61.

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing:

    Mapping the territories. *Applied Linguistics*, *38*(2), 165-193. DOI:

    10.1093/applin/amv011

Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use

    of collocations?. *International Review of Applied Linguistics in Language*

Egbert, J., & Staples, S. (2019). Doing Multi-Dimensional Analysis in SPSS, SAS, and R. In

    T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional Analysis: Research Methods*

    and Current Issues, (pp. 27-42). Bloomsbury.

Elbow, P. (1991). Reflections on academic discourse: How it relates to freshmen and

    colleagues. *College English*, *53*(2), 135-155.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*

    20(1), 29–62. doi: 10.1515/text.1.2000.20.1.29

Flowerdew, J. (2013a). Some thoughts on English for research publication purposes (ERPP) and related issues. *Language Teaching*, *48*(2), 250-262. DOI: 10.1017/S0261444812000523

Flowerdew, J. (2013b). English for research publication purposes. In B. Paltridge and S. Starfield (Eds), *Handbook of English for Specific Purposes* (pp. 301–321). Malden, MA: Wiley-Blackwell.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp. 75-94). Harlow: Longman.

Friginal, E & Hardy, (2019). From Factors to Dimensions: Interpreting Linguistic Co-occurrence Patterns. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional Analysis: Research Methods and Current Issues* (pp. 145-164). Bloomsbury.

Fuller, S. (1991). Disciplinary boundaries and the rhetoric of the social sciences. *Poetics Today*, *12*(2), 301-325.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics, 28*(2), 241-265. DOI: 10.1093/applin/amm010

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305-327.

Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*. Advance online publication. DOI: 10.1093/applin/amy005

Gray, B & Biber, D. (2015). Stance Markers. In K. Aihmer & C. Ruhlemann (Eds.), *Corpus of Pragmatics* (pp. 219 -248). Cambridge University Press.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam, Netherlands: John Benjamins.

Gray, B., Cotos, E., & Smith, J. (2020). Combining rhetorical move analysis with multi-dimensional analysis. In U. Romer, V. Cortes, & E. Friginal (Eds.), *Advances in Corpus-based Research on Academic Writing: Effects of discipline, register, and writer expertise* (pp. 137 – 168). John Benjamins.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics 13*(4), 403–37.

Gries, S. T. (2013). *Statistics for linguistics with R: A practical introduction*. Walter de Gruyter.

Gries, S. Th. (2015). The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora*, *10*(1), 95-125. DOI: 10.3366/cor.2015.0068

Gries, S. Th. & Durrant, P. (2019). Analyzing co-occurrence data. In M. Paquot & S. Th. Gries (Eds.). *Practical handbook of corpus linguistics* (pp. 1-15). Berlin & New York: Springer

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes, 4*, 257-277.

Groom, N. (2009). Phraseology and epistemology in academic book reviews: A corpus-driven analysis of two humanities disciplines. In K. Hyland and D. Giuliana (Eds.) *Academic Evaluation. Review Genres in University Settings* (pp. 122–139). Basingstoke: Palgrave Macmillan.

Halliday, M.A.K. (1989). *Spoken and written language*. Oxford University Press.

Halliday, M.A.K. (2004). *The Language of Science*. London: Continuum.

Haswell, R. (1991). *Gaining ground in college writing: Tales of development and interpretation*. Southern Methodist University Press.

Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general

    academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to*

    say about vocabulary instruction (pp. 106–129). Newark, DE: International Reading

    Association.

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.

Hopkins, A., & Dudley-Evans, T. (1988). A genre-based investigation of the discussion

    sections in articles and dissertations. *English for Specific Purposes, 7*, 113-122.

Hoskin, K. (1993). Education and the genesis of disciplinarity: the unexpected reversal. In E.

    Messer-davidow, D. R. Shumway, & D. J. Sylvan (Eds.). *Knowledges: historical and*

    critical studies in disciplinarity (pp. 271-304). University of Virginia Press.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.),

    Phraseology: Theory, analysis, and application (pp. 161-186). Oxford: Oxford

    University.

    http://www.stm assoc.org/2009_10_13_MWC_STM_Report.pdf

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical*

    grammar of English. John Benjamins Publishing.

Hutchinson, T. & Waters, A. (1987). *English for Specific Purposes*. Cambridge: Cambridge

    University Press.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. Applied

    Linguistics 17(4): 433–454. DOI: 10.1093/applin/17.4.433

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *Text-*

    Interdisciplinary Journal for the Study of Discourse, *18*(3), 349-382.

Hyland, K. (2001). Bringing in the reader: Addressee features in academic articles. *Written communication*, *18*(4), 549-574.

Hyland, K. (2002a). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115–130). New York: Longman.

Hyland, K. (2002b). Directives: Argument and engagement in academic writing. *Applied Linguistics, 23*, 215–239.

Hyland, K. (2004a). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.

Hyland, K. (2004b). Graduates' aratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes 23*(3), 303-324.

Hyland, K. (2004c). *Disciplinary discourses*. Ann Arbor: University of Michigan Press.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173-192. DOI: 10.1177/1461445605050365

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21. DOI: 10.1016/j.esp.2007.06.001

Hyland, K. (2011). Disciplines and discourses: Social interactions in the construction of knowledge. In D. Starke-Meyerring, A. Pare, N. Artemeva, M. Horne, & L. Yousoubova (Eds.), *Writing in knowledge societies* (pp. 193-214). Fort Collins, CO: The WAC Clearinghouse and Parlor Press.

Hyland, K. (2013). ESP and writing. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 95-113). West Sussex, UK: John Wiley & Sons.

Hyland, K. (2015). Genre, discipline and identity. *Journal of English for Academic Purposes*, *19*, 32-43. DOI: 10.1016/j.jeap.2015.02.005

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing, 31*, 58-69. DOI: 10.1016/j.jslw.2016.01.005

Hyland, K. (2018). *The Essential Hyland: Studies in Applied Linguistics*. Bloomsbury Publishing.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for specific purposes*, *24*(2), 123-139.

Hyland, K., & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL quarterly*, *41*(2), 235-253. DOI: 10.1002/j.1545-7249.2007.tb00058.x

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Jesse, E. (2015). Publication type and discipline variation in published academic writing. *International Journal of Corpus Linguistics*, *20*(1), 1-29.

Jiang, F. K., & Hyland, K. (2017). Metadiscursive nouns: Interaction and cohesion in abstract moves. *English for Specific Purposes*, *46*, 1-14. DOI: 10.1016/j.esp.2016.11.001

Jiang, F., & Hyland, K. (2015). 'The fact that': Stance nouns in disciplinary writing. *Discourse Studies*, *17*(5), 529-550. DOI: 10.1177/1461445615590719

Jordan, R. (1997). *English for Academic Purposes*. Cambridge: Cambridge University Press.

Kaiser, H. (1974). An Index of Factorial Simplicity. *Psychometrika, 39* (1), 31–36.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, *24*, 269–292.

Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In D. Biber, U. Connor, & T. Upton (Eds.), *Discourse on the move: Using corpus analysis to describe discourse structure* (pp. 73–119). John Benjamins.

Keller, F., & Lapata, M. (2003). Using the web to obtain frequencies for unseen bigrams. *Computational linguistics, 29*(3), 459-484.

Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Longman.

Kolb, D. A. (1976). Management and the learning process. *California management review*, *18*(3), 21-31.

Kolb, D. A. (1981). Learning styles and disciplinary differences. In A. W. Chickering (Ed.), *The modern American college* (pp. 232-255). Jossey-Bass.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. University of Chicago press.

Layder, D. (1993). *New Strategies in Social Research*. Polity Press.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes, 39*, 45-61. DOI: 10.1016/j.esp.2015.03.003

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*(2), 85-102.

Lijffijt, J., & Gries, S. T. (2012). Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora". *International Journal of Corpus Linguistics 17*(a), 147-149.

Lillis, T. & Curry, M.J. (2010). A*cademic Writing in a Global Context: The Politics and Practices of Publishing in English*. Routledge.

Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, *25*(3), 282-309. DOI: 10.1016/j.esp.2005.07.001

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, *31*(3), 150-160.

Lin, P. M. (2013). The prosody of formulaic expression in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics, 18*(4), 561-588.

Lin, P. M. (2018). Lin, P. (2018). *The Prosody of Formulaic Sequences: A Corpus and Discourse Approach*. New York/London: Bloomsbury Publishing.

Liu, D. & Lei, L. (2019). Technical vocabulary. In S. Webb (Ed), *The Routledge handbook of vocabulary studies* (pp. 111-124). Routledge.

Lu, X., Casal, J. E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes*, *44*, 100832.

Mannheim, K. (2013). *Ideology and utopia*. Routledge.

Manning, C. & Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Martinez, 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. DOI: 10.1016/j.esp.2009.04.003

Martinez, R. (2018). "Specially in the last years…": Evidence of ELF and non-native English forms in international journals. *Journal of English for Academic Purposes*, *33*, 40-52.

Mauranen, A., Hynninen, N., & Ranta, E. (2016). English as the academic lingua franca. In K. Hyland, & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 44-55). New York: Routledge.

McEnery, T, Xiao, R & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. New York NY: Routledge.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics 32*, 111–129. DOI: 10.1017/S0267190512000128

Mizumoto, A., Hamatani, S., & Imao, Y. (2017). Applying the Bundle–Move Connection Approach to the Development of an Online Writing Support Tool for Research Articles. *Language Learning*. DOI: 10.1111/lang.12250

Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford, UK: Clarendon Press.

Moore, T. (2002). Knowledge and Agency: A Study of 'Metaphenomenal Discourse' in Textbooks from Three Disciplines. *English for Specific Purposes 21*(4), 347-366. DOI: 10.1016/S0889-4906(01)00030-8.

Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, *50*, 40-63. DOI: 10.1016/j.esp.2017.11.006

Nation, I. S. (2013). *Teaching & learning vocabulary*. Heinle Cengage Learning.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, P., & Coxhead, A. (2012). Special purposes vocabulary. In C. A. Chapelle (Ed), *The Encyclopedia of Applied Linguistics* (pp. 2-4). Blackwell Publishing Ltd.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, England: Oxford University Press.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education. Cambridge:* Cambridge University Press.

Nesi, H., Gardner, S., Thompson, P. & Wickens, P. (2008-2010). *The British Academic Written English (BAWE) corpus.*

Nesselhauf, N. (2004). *Collocations in a learner corpus.* John Benjamins. doi: 10.1075/scl.14

Omidian, Siyanova-Chanturia, & Durrant (2020). Predicting parameters of variation in the use of academic formulas in university student writing. In P. Szudarski & S. Barclay (eds), *Vocabulary Theory, Patterning and Teaching*. Clevedon: Multilingual Matters.

Omidian, T., Shahriari, H., & Siyanova-Chanturia, A. (2018). A cross-disciplinary investigation of multi-word expressions in the moves of research article abstracts. *Journal of English for academic purposes*, *36*, 1-14. DOI: 10.1016/j.jeap.2018.08.002

Ozturk, I. (2007). The textual organization of research article introductions in applied

>   linguistics: Variability within a single discipline. *English for Specific Purposes 26*().

>   25–38. DOI: 10.1016/j.esp.2005.12.003

Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. Applied

>   Linguistics, 15, 288-299. DOI: 10.1093/applin/15.3.288

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English

>   academic professionals: Lexical bundles in Telecommunications research

>   journals. *Journal of English for Academic Purposes*, *21*, 60-71. DOI:

>   10.1016/j.jeap.2015.11.003

Pantin, C. F. A. (1968). *Relations between sciences*. Cambridge University Press.

Paquot, M. (2007). Towards a productively-oriented academic word list. In J. Walinski, K.

>   Kredens, & S. Gozdz-Roszkowski (Eds.), Corpora and ICT in language studies.

>   PALC 2005. Lodz studies in LANGUAGE 13 (pp. 127–140). Frankfurt am main:

>   Peter Lang.

Parkinson, J. (2013). Representing own and other voices in social science research

>   articles. *International Journal of Corpus Linguistics*, *18*(2), 199-228. DOI:

>   10.1075/ijcl.18.2.02par

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: Theory

>   and practice in corpus-assisted discourse studies*. John Benjamins Publishing.

Pe´rez-Llantada, C. (2014). *Scientific discourse and the rhetoric of globalization: The impact

>   of culture and language*. London: Bloomsbury.

Peacock, M. (2002). Communicative moves in the discussion section of research

>   articles. *System*, *30*(4), 479-497.

Peters, A. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University

>   Press.

Post, R. (2009). Debating disciplinarity. *Critical inquiry*, *35*(4), 749-770.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519-549. DOI: 10.1075/ijcl.13.4.06ray

Rayson, P. & Garside, R. (2000). Comparing corpora using frequency profiling. In proceedings of the *workshop on Comparing Corpora, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000)*. Hong Kong.

Rayson, P., & Garside, R. (2000, October). Comparing corpora using frequency profiling. In *Proceedings of the workshop on Comparing corpora-Volume 9* (pp. 1-6). Association for Computational Linguistics.

Rea Rizzo, C. (2010). Getting on with corpus compilation: From theory to practice. *ESP World*, *1*(9), 1-22.

Romer, U. (2009). English in academia: does nativeness matter? *Anglistik: International Journal of English Studies*, *20*(2), 89-100.

Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English text construction*, *3*(1), 95-119.

Russell, D. (1991). *Writing in the academic disciplines: A curricular history.* Southern Illinois University Press.

Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, *21*(1), 1-17. DOI: 10.1016/S0889-4906(00)00023-5

Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes*, *24*(2), 141-156. 10.1016/j.esp.2002.10.001.

Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, *7*(1), 55-67. DOI: 10.1016/j.jeap.2008.02.005

Samraj, B. (2016). Discourse structure and variation in manuscript reviews: Implications for genre categorization. *English for Specific Purposes*, *42*, 76-88. DOI: 10.1016/j.esp.2015.12.003

Samraj, B., & Monk, L. (2008). The statement of purpose in graduate program applications: Genre structure and disciplinary variation. *English for Specific Purposes*, *27*(2), 193-211. DOI: 10.1016/j.esp.2007.07.001

Sardinha & M. V. Pinto (2019). *Multi-dimensional Analysis: Research Methods and Current Issues*. Bloomsbury.

Sardinha, T. B., Pinto, M. V., Mayer, C., Zuppardi, M. C., & Kauffmann, C. H. (2019). Adding registers to a previous Multi-dimensional Analysis. In T. B. Sardinha & M. V. Pinto (Eds.), *Multi-dimensional Analysis: Research Methods and Current Issues* (pp. 165-186). Bloomsbury.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson, and E. Judd (Eds.), *Sociolinguistics and Language Acquisition* (pp. 137-174). Rowley, MA: Newbury House.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913-951. DOI: 10.1111/lang.12077

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1-23). Amsterdam: John Benjamins. doi: 10.1075/lllt.9.02sch

Scott, M. (1996). *WordSmith Tools.* Oxford: Oxford University Press.

Scott, M. & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins Publishing.

Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.

Shulman, L. (1981). Disciplines of inquiry in education: An overview. *Educational researcher*, *10*(6), 5-23.

Shumway, D. R. & Messer-Davidow, E. (1991). Disciplinarity: an introduction. *Poetics today, 12*(2), 201-225.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press

Siyanova-Chanturia, A & Omidian, T. (2020). Key issues in research multi-word expressions. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 529-544)*.* London: Routledge.

Siyanova-Chanturia, A. & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 38-61). New York and London: Routledge.

Siyanova, A. & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64*(3), 429–458.

Sorhus, H. B. (1977). To Hear Ourselves--Implications for Teaching English as a Second Language. *English Language Teaching Journal*, *31*(3), 211-221.

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. DOI: 10.1016/j.esp.2012.09.001

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.

Straus, R. (1973). Departments and Disciplines: Stasis and Change: Significant multidisciplinary activity requires a redistribution of power within universities. *Science*, *182*(4115), 895-898.

Swales, J. (1981). *Aspects of article introductions*. Birmingham, UK: The University of Aston, Language Studies Unit.

Swales, J. (1984). Research into the structure of introductions to journal articles and its application to the teaching of academic writing. In R. Williams & J. Kirkman (Eds.), *Common Grounds: Shared interests in ESP and communication studies* (pp. 77–86). New York NY: Pergamon Press.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swales, J. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press.

Swales, J. M., & Feak, C. B. (2009). *Abstracts and the writing of abstracts*. University of Michigan Press.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students*. University of Michigan Press.

Tabachnick, B. G. & Fidell, L. S. (2007), *Using Multivariate Statistics*. Pearson Education.

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic

    papers. *Applied linguistics*, *12*(4), 365-382.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Van Lancker-Sidtis D. & Rallon, G. (2004). Tracking the incidence of formulaic expressions

    in everyday speech: Methods for classification and verification. *Language &*

    Communication 24(3), 207–240. DOI: 10.1016/j.langcom.2004.02.003

Vihman, M. (1982). The acquisition of morphology by a bilingual child: a whole-word

    approach. *Applied Psycholinguistics 3*, 141–60.

Von Eye, A., Mun, E. Y., & Mair, P. (2012). Log-linear modeling. *Wiley Interdisciplinary*

    Reviews: Computational Statistics, *4*(2), 218-223.

Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word

    list. *English for Specific Purposes*, *27*(4), 442-458. DOI: 10.1016/j.esp.2008.05.003

Ware, M., & Mabe, M. (2009). The STM report: An overview of scientific and scholarly

    journal publishing. Retrieved from

West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: a practical guide*

    using statistical software. Chapman and Hall/CRC.

Wood, A. (2001). International scientific English: the language of research scientists around

    the world. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English*

    for academic purposes (pp. 71-83). Cambridge: Cambridge University Press.

Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching.

    TESL Canada Journal 20(1), 01-15.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, England: Cambridge

    University Press.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated

    model. *Language & Communication*, *20*(1), 1-28.

Yang, R. & Allison, D. (2003). RAs in applied linguistics: Moving from results to

    conclusions. English for Specific Purposes, 22(4), pp.365-384.

Zipf, G.K. (1949). *Human behavior and the principle of least effort*. Cambridge, Mass:

    Addison-Wesley.

Appendix A

Overlaps between sections across disciplines

*I: Introduction M: Methods R: Results D: Discussion
*BI: Biology; CH: Chemistry; DE: Dentistry; ME: Mechanical Engineering; PH: Physics; AL: Applied Linguistics; BU: Business; MA: Management; PO: Politics; SO: Sociology

	bi.i	bi.m	bi.r	bi.d	ch.i	ch.m	ch.r	ch.d	de.i	de.m	de.r	de.d	me.i	me.m	me.r	me.d	ph.i	ph.m	ph.r	ph.d	ap.i	ap.m	ap.r	ap.d	bu.i	bu.m	bu.r	bu.d	ma.i	ma.m	ma.r	ma.d	po.i	po.m	po.r	po.d	so.i	so.m	so.r	so.d
bi.i	--	26.02	40.25	60.49	40.92	15.34	26.72	33.19	25.36	12.81	17.2	26.04	24.79	17.14	17.81	19.25	22.87	17.36	16.14	23.66	20.65	14.69	18.03	22.13	20.1	17.17	20.04	20.22	23.66	19.47	25.12	23.11	18.43	18.76	16.98	23.67	25.6	22.22	17.69	29.33
bi.m	26.02	--	43.44	32.08	21.25	48.57	36.17	30.25	11.6	38.12	18.29	25.07	21.51	31.36	29.97	28.64	12.83	31.66	27.3	23.51	10.47	23.47	20.43	15.16	9.79	28.07	22.22	13.1	10.01	28.17	24.45	15.34	5.41	21.3	20.43	10.95	10.14	26.98	20.69	14.24
bi.r	40.25	43.44	--	59.89	29.71	30.79	53.95	48.22	19.33	26.66	31.38	40.48	25.81	25.35	36.96	35.61	18.5	24.95	33.18	32.81	15.5	18.42	27.65	22.32	15.5	24.7	29.82	20.23	17.49	25.27	29.89	23.08	10.43	21.69	31.01	30	18.9	26.12	32.33	27.09
bi.d	60.49	32.08	59.89	--	34.72	18.88	39.07	48.74	25.18	16.6	25.49	41.14	26	18.94	30.33	31.34	23.9	19.34	27	33.92	21.73	15.91	27.54	29.82	22.93	23.58	31.1	29.09	26.03	26.36	34.25	34	18.95	23.59	30	33.46	26.69	26.92	30.81	36.26
ch.i	40.92	21.25	29.71	34.72	--	30.09	41.6	51.85	37.73	19.24	20.89	33.89	35.55	22	22.63	25.3	29.83	20.8	16.47	25.23	19.04	13.69	16.43	18.02	16.8	14.42	14.41	16.76	23.25	16.59	17.86	17.85	18.43	14.21	12.17	18.65	23.34	16.11	11.78	21.86
ch.m	15.34	48.57	30.79	18.88	30.09	--	42.27	35.72	11.85	44.52	20.09	25.4	20.8	32.05	24.74	24.45	16.09	29.62	24.03	20	8.85	19.03	13.77	10.69	8.85	20.21	15.05	9.49	7.85	22.5	17.67	10.28	6.63	15.35	12.66	9.88	9.19	16.66	11.71	9.67
ch.r	26.72	36.17	53.95	39.07	41.6	42.27	--	64.06	24.74	28.72	32.58	41.81	28.92	31.11	41.79	39.74	22.06	29.76	38.79	37.74	13.93	16.87	22.22	19.73	12.7	21.42	24.78	17.7	14.31	23.49	27.71	19.97	8.72	16.76	24.02	16.63	14.28	20.48	25	20.35
ch.d	33.19	30.25	48.22	48.74	51.85	35.72	64.06	--	25	24.28	27.58	46.75	29.98	26.04	38.24	40.6	24.24	24.56	33.38	38.84	17.55	14.21	22.22	23.91	17.93	21.47	25.92	23.82	20.74	23.42	27.71	25.3	12.47	18.53	24.88	23.14	18.18	20.27	23.96	25.74
de.i	25.36	11.6	19.33	25.18	37.73	11.85	24.74	25	--	29.57	34.85	54.08	26.42	16.75	16.4	17.83	17.91	11.28	10.65	18.82	15.15	7.79	13.55	14.22	13.33	13.58	15.31	13.58	16.4	15.38	17.02	14.63	15.68	15.25	12	15.04	21.37	15.31	12.86	17.96
de.m	12.81	38.12	26.66	16.6	19.24	44.52	28.72	24.28	29.57	--	37.81	41.22	23.07	32.78	24.36	24.68	12.58	24.58	19.76	18.05	12.21	24.82	20.22	13.74	13.03	24.42	21.28	12.98	11.71	29.54	25.32	14.69	8.17	23.1	21.54	11.53	12.86	24.27	18.35	14.65
de.r	17.2	18.29	31.38	25.49	20.89	20.09	32.58	27.58	34.85	37.81	--	45.56	20.23	20.25	22.52	22.78	18.45	22.36	22.35	25.75	11.41	15.45	22.73	16.77	12.61	18.4	23.5	13.9	11.87	20.08	28.49	14.17	10.07	21.08	27.52	15.46	17.74	20	23.5	19.27
de.d	26.04	25.07	40.48	41.14	33.89	25.4	41.81	46.75	54.08	41.22	45.56	--	33.26	26.86	33.03	37.44	22.27	21.56	26.32	33.63	21.07	17.69	23.64	25.16	20.66	23.43	29.74	25	21.23	25.61	29.81	25.32	14.18	21.49	26.57	24.95	23.87	23.89	27.67	30.03
me.i	24.79	21.51	25.81	26	35.55	20.8	28.92	29.98	26.42	23.07	20.23	33.26	--	59.11	52.71	57.95	31.4	23.67	23.92	31.55	20.7	19.76	17.28	21.65	17.88	18.93	18.25	16.52	19.9	24	21.65	18.47	10.85	16.45	12.87	15.85	18.18	17.75	14.21	18.93
me.m	17.14	31.36	25.35	18.94	22	32.05	31.11	26.04	16.75	32.78	20.25	26.86	59.11	--	57.31	58.21	22.33	34.12	29.18	28.67	12.65	20.54	15.9	15.15	14.34	23.17	18.43	13.33	10.41	23.7	23.07	12.99	8.02	18.78	14.55	9.94	11.52	19.96	15.65	12.94
me.r	17.81	29.97	36.96	30.33	22.63	24.74	41.79	38.24	16.4	24.36	22.52	33.03	52.71	57.31	--	75.7	22.5	30.69	44.81	41.07	14.23	15.76	23.59	20.71	12.25	21.31	25.53	17.1	12.18	22.77	28.61	19.08	7.18	17.16	23.19	16.42	13.47	20.79	24.86	19.24
me.d	19.25	28.64	35.61	31.34	25.3	24.45	39.74	40.6	17.83	24.68	22.78	37.44	57.95	58.21	75.7	--	23.15	30.79	38.37	39.61	16.34	16.45	20.99	21.37	14.56	22.53	24.69	18.63	14.54	25	28.24	21.93	8.2	17	23.72	17.56	16.06	20.86	24.33	21.08
ph.i	22.87	12.83	18.5	23.9	29.83	16.09	22.06	24.24	17.91	12.58	18.45	22.27	31.4	22.33	22.5	23.15	--	44.69	39.1	56.73	17.22	12.24	15.13	16.66	15.55	14.72	17.35	13.95	19.59	13.6	17.73	13.85	16.14	16.12	15.98	19.53	21.25	15.6	14.06	19
ph.m	17.36	31.66	24.95	19.34	20.8	29.62	29.76	24.56	11.28	24.58	22.36	21.56	23.67	34.12	30.69	30.79	44.69	--	51.02	51.89	10.7	18.29	16.99	13.55	9.73	19.04	19.14	10.93	12.06	18.28	22.75	12.37	9.52	16.96	18.28	13.18	11.32	20.81	17.4	12.17
ph.r	16.14	27.3	33.18	27	16.47	24.03	38.79	33.38	10.65	19.76	22.35	26.32	23.92	29.18	44.81	38.37	39.1	51.02	--	66.16	11.68	14.41	20.78	17.8	11.68	22.51	26.57	16.68	13.31	21.27	28.52	17.2	6.87	17.92	26.21	16.87	15.38	23.41	28.22	20.26
ph.d	23.66	23.51	32.81	33.92	25.23	20	37.74	38.84	18.82	18.05	25.75	33.63	31.55	28.67	41.07	39.61	56.73	51.89	66.16	--	17.31	15.19	21.37	19.5	14.43	19.96	25.59	18.72	16.94	20.98	27.49	19.47	11.21	17.62	25.17	21.4	20.22	23.2	25.11	23.63
ap.i	20.65	10.47	15.5	21.73	19.04	8.85	13.93	17.55	15.15	12.21	11.41	21.07	20.7	12.65	14.23	16.34	17.22	10.7	11.68	17.31	--	50.89	49.29	61.62	31.75	22.83	22.94	29.06	36.18	24.13	22.64	29.76	20.74	21.78	20.19	25.27	32.98	22.95	18.11	30.15
ap.m	14.69	23.47	18.42	15.91	13.69	19.03	16.87	14.21	7.79	24.82	15.45	17.69	19.76	20.54	15.76	16.45	12.24	18.29	14.41	15.19	50.89	--	58.91	48.47	20.27	38.54	24.5	20.63	22.04	40.44	26.95	22.66	14.01	30	21.12	17.29	21.16	33.33	20.76	22.22
ap.r	18.03	20.43	27.65	27.54	16.43	13.77	22.22	22.22	13.55	20.22	22.73	23.64	17.28	15.9	23.59	20.99	15.13	16.99	20.78	21.37	49.29	58.91	--	60.23	22.69	32.77	33.68	27.77	21.41	33.67	35.73	28.68	12.67	28.28	32.98	24.62	23.28	33.86	32.68	29.1
ap.d	22.13	15.16	22.32	29.82	18.02	10.69	19.73	23.91	14.22	13.74	16.77	25.16	21.65	15.15	20.71	21.37	16.66	13.55	17.8	19.5	61.62	48.47	60.23	--	35.79	32.09	35.45	41.26	34.4	32.68	34.35	42.95	19.27	24.17	27.9	32.22	30.27	31.39	30.83	39.1
bu.i	20.1	9.79	15.5	22.93	16.8	8.85	12.7	17.93	13.33	13.03	12.61	20.66	17.88	14.34	12.25	14.56	15.55	9.73	11.68	14.43	31.75	20.27	22.69	35.79	--	45.32	49.33	65.05	61.61	39.12	35.04	47.36	35.15	27.45	27.04	39.91	39.79	27.32	23.34	40.47
bu.m	17.17	28.07	24.7	23.58	14.42	20.21	21.42	21.47	13.58	24.42	18.4	23.43	18.93	23.17	21.31	22.53	14.72	19.04	22.51	19.96	22.83	38.54	32.77	32.09	45.32	--	60.38	50.13	36.1	67.14	44.87	40.71	16.3	43.25	34.36	25.7	25.27	49.36	34.79	32.72
bu.r	20.04	22.22	29.82	31.1	14.41	15.05	24.78	25.92	15.31	21.28	23.5	29.74	18.25	18.43	25.53	24.69	17.35	19.14	26.57	25.59	22.94	24.5	33.68	35.45	49.33	60.38	--	58.61	38.03	49.07	56.59	45.73	20.53	35.35	44.72	34.05	29.39	40.92	46.51	39.33
bu.d	20.22	13.1	20.23	29.09	16.76	9.49	17.7	23.82	13.58	12.98	13.9	25	16.52	13.33	17.1	18.63	13.95	10.93	16.68	18.72	29.06	20.63	27.77	41.26	65.05	50.13	58.61	--	50.26	42.67	40.06	62.99	24.65	27.31	33.77	39.86	33.45	30.07	33.42	46.36
ma.i	23.66	10.01	17.49	26.03	23.25	7.85	14.31	20.74	16.4	11.71	11.87	21.23	19.9	10.41	12.18	14.54	19.59	12.06	13.31	16.94	36.18	22.04	21.41	34.4	61.61	36.1	38.03	50.26	--	42.32	39.56	57.51	31.73	23.76	23.82	36.98	42.27	27.61	23.52	40.73
ma.m	19.47	28.17	25.27	26.36	16.59	22.5	23.49	23.42	15.38	29.54	20.08	25.61	24	23.7	22.77	25	13.6	18.28	21.27	20.98	24.13	40.44	33.67	32.68	39.12	67.14	49.07	42.67	42.32	--	56.32	46.93	20.33	45.2	38.15	27.08	29.58	52.22	37.19	34.97
ma.r	25.12	24.45	29.89	34.25	17.86	17.67	27.71	27.71	17.02	25.32	28.49	29.81	21.65	23.07	28.61	28.24	17.73	22.75	28.52	27.49	22.64	26.95	35.73	34.35	35.04	44.87	56.59	40.06	39.56	56.32	--	46.79	18.82	35.64	44.83	31.38	25.7	41.68	43.54	34.54
ma.d	23.11	15.34	23.08	34	17.85	10.28	19.97	25.3	14.63	14.69	14.17	25.32	18.47	12.99	19.08	21.93	13.85	12.37	17.2	19.47	29.76	22.66	28.68	42.95	47.36	40.71	45.73	62.99	57.51	46.93	46.79	--	21.09	26.58	32.41	38.53	30.08	31.91	33.46	42.71
po.i	18.43	5.41	10.43	18.95	18.43	6.63	8.72	12.47	15.68	8.17	10.07	14.18	10.85	8.02	7.18	8.2	16.14	9.52	6.87	11.21	20.74	14.01	12.67	19.27	35.15	16.3	20.53	24.65	31.73	20.33	18.82	21.09	--	44.27	38.66	62.76	39.73	18.98	17.23	31.7
po.m	18.76	21.3	21.69	23.59	14.21	15.35	16.76	18.53	15.25	23.1	21.08	21.49	16.45	18.78	17.16	17	16.12	16.96	17.92	17.62	21.78	30	28.28	24.17	27.45	43.25	35.35	27.31	23.76	45.2	35.64	26.58	44.27	--	55.51	47.54	31.5	49.48	34.04	33.27
po.r	16.98	20.43	31.01	30	12.17	12.66	24.02	24.88	12	21.54	27.52	26.57	12.87	14.55	23.19	23.72	15.98	18.28	26.21	25.17	20.19	21.12	32.98	27.9	27.04	34.36	44.72	33.77	23.82	38.15	44.83	32.41	38.66	55.51	--	53.42	29.27	40.18	49.92	39.2
po.d	23.67	10.95	21.28	33.46	18.65	9.88	16.63	23.14	15.04	11.53	15.46	24.95	15.85	9.94	16.42	17.56	19.53	13.18	16.87	21.4	25.27	17.29	24.62	32.22	39.91	25.7	34.05	39.86	36.98	27.08	31.38	38.53	62.76	47.54	53.42	--	38.92	27.33	31.84	46.9
so.i	25.6	10.14	18.9	26.69	23.34	9.19	14.28	18.18	21.37	12.86	17.74	23.87	18.18	11.52	13.47	16.06	21.25	11.32	15.38	20.22	32.98	21.16	23.28	30.27	39.79	25.27	29.39	33.45	42.27	29.58	25.7	30.08	39.73	31.5	29.27	38.92	--	40.86	42.32	64.22
so.m	22.22	26.98	26.12	26.92	16.11	16.66	20.48	20.27	15.31	24.27	20	23.89	17.75	19.96	20.79	20.86	15.6	20.81	23.41	23.2	22.95	33.33	33.86	31.39	27.32	49.36	40.92	30.07	27.61	52.22	41.68	31.91	18.98	49.48	40.18	27.33	40.86	--	56.49	48.81
so.r	17.69	20.69	32.33	30.81	11.78	11.71	25	23.96	12.86	18.35	23.5	27.67	14.21	15.65	24.86	24.33	14.06	17.4	28.22	25.11	18.11	20.76	32.68	30.83	23.34	34.79	46.51	33.42	23.52	37.19	43.54	33.46	17.23	34.04	49.92	31.84	42.32	56.49	--	59.45
so.d	29.33	14.24	27.09	36.26	21.86	9.67	20.35	25.74	17.96	14.65	19.27	30.03	18.93	12.94	19.24	21.08	19	12.17	20.26	23.63	30.15	22.22	29.1	39.1	40.47	32.72	39.33	46.36	40.73	34.97	34.54	42.71	31.7	33.27	39.2	46.9	64.22	48.81	59.45	--

Appendix B

All the bundles identified in the DCRA, their functional and structural categories, and their distinctiveness

BUNDLES	FUN	STR	INTRODUCTION				METHODS				RESULTS				DISCUSSION			
			I.H	I.S	G2	SIG.	M.H	M.S	G2	SIG.	R.H	R.S	G2	SIG.	D.H	D.S	G2	SIG.
A BETTER UNDERSTANDING OF	T	NP	32.4	56.5	1.6	not	0.0	6.7	3.6	not	2.2	7.4	1.4	not	9.9	50.6	**12.9**	**<.001**
A BETTER UNDERSTANDING OF THE	T	NP	13.9	22.6	0.5	not	0.0	4.5	2.4	not	2.2	3.7	0.2	not	9.9	19.0	1.3	not
A BROAD RANGE OF	R	NP	13.9	11.3	0.1	not	5.3	13.4	1.4	not	4.4	3.7	0.0	not	9.9	10.5	0.0	not
A CHANGE IN THE	R	NP	13.9	11.3	0.1	not	8.0	8.9	0.0	not	17.6	1.9	**7.5**	**<.01**	17.3	6.3	2.3	not
A COMPARISON OF THE	T	NP	0.0	11.3	3.6	not	2.7	6.7	0.7	not	28.7	11.1	**4.0**	**<.05**	14.8	4.2	2.8	not
A CONSEQUENCE OF THE	T	NP	4.6	7.5	0.2	not	5.3	4.5	0.0	not	13.2	9.3	0.3	not	12.3	14.8	0.1	not
A CRITICAL ROLE IN	P	NP	18.5	3.8	2.6	not	0.0	2.2	1.2	not	6.6	0.0	**4.7**	**<.05**	17.3	6.3	2.3	not
A DECREASE IN THE	R	NP	18.5	0.0	**6.4**	**<.05**	0.0	0.0	0.0	not	22.1	5.6	**5.3**	**<.05**	22.2	0.0	**14.0**	**<.001**
A DEEPER UNDERSTANDING OF	T	NP	4.6	15.1	1.4	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	2.5	10.5	2.3	not
A DEEPER UNDERSTANDING OF THE	T	NP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	2.2	1.9	0.0	not	2.5	10.5	2.3	not
A FINAL CONCENTRATION OF	R	NP	0.0	0.0	0.0	not	50.6	0.0	**29.9**	**<0.0001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
A FUNCTION OF THE	T/R	NP	41.6	3.8	**9.1**	**<.01**	34.6	22.3	1.1	not	75.0	18.6	**18.3**	**<0.0001**	42.0	10.5	**9.0**	**<.01**
A GREAT DEAL OF	R	NP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	0.0	11.1	**7.3**	**<.01**	7.4	23.2	3.7	not
A GREATER NUMBER OF	R	NP	0.0	26.4	**8.3**	**<.01**	0.0	6.7	3.6	not	4.4	5.6	0.1	not	14.8	6.3	1.6	not
A GROWING BODY OF	T	NP	0.0	33.9	**10.7**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	10.5	2.3	not
A GROWING BODY OF RESEARCH	T	NP	0.0	22.6	**7.1**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
A HIGH DEGREE OF	R	NP	4.6	3.8	0.0	not	0.0	13.4	**7.3**	**<.01**	8.8	7.4	0.1	not	9.9	19.0	1.3	not
A HIGH LEVEL OF	R	NP	9.3	18.8	0.8	not	2.7	24.5	**8.1**	**<.01**	4.4	29.7	**10.1**	**<.01**	4.9	52.7	**19.7**	**<0.0001**
A HIGHER LEVEL OF	R	NP	4.6	0.0	1.6	not	0.0	8.9	**4.9**	**<.05**	0.0	11.1	**7.3**	**<.01**	9.9	19.0	1.3	not
A LARGE BODY OF	R	NP	4.6	30.1	**4.9**	**<.05**	0.0	0.0	0.0	not	0.0	1.9	1.2	not	0.0	8.4	**4.9**	**<.05**
A LARGE BODY OF RESEARCH	R	NP	4.6	22.6	3.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
A LARGE BODY OF RESEARCH HAS	T	NP_VP	4.6	15.1	1.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
A LARGE NUMBER OF	R	NP	23.1	22.6	0.0	not	13.3	20.0	0.6	not	28.7	16.7	1.6	not	24.7	16.9	0.7	not
A LARGER NUMBER OF	R	NP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	0.0	9.3	**6.1**	**<.05**	4.9	14.8	2.2	not
A LIMITED NUMBER OF	R	NP	18.5	11.3	0.4	not	8.0	4.5	0.4	not	4.4	1.9	0.5	not	4.9	14.8	2.2	not
A LOW LEVEL OF	R	NP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	4.4	14.9	2.9	not	0.0	10.5	**6.2**	**<.05**
A LOWER LEVEL OF	R	NP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	4.4	11.1	1.5	not	0.0	2.1	1.2	not
A MEASURE OF THE	R	NP	0.0	0.0	0.0	not	21.3	22.3	0.0	not	6.6	9.3	0.2	not	2.5	4.2	0.2	not
A MORE COMPREHENSIVE UNDERSTANDING OF	T	NP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	14.8	**8.6**	**<.01**
A NEGATIVE EFFECT ON	R	NP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	4.4	16.7	3.7	not	0.0	4.2	2.5	not
A NUMBER OF STUDIES	T	NP	41.6	22.6	1.4	not	0.0	4.5	2.4	not	0.0	0.0	0.0	not	7.4	8.4	0.0	not
A NUMBER OF STUDIES HAVE	T	NP	27.8	15.1	0.9	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
A NUMBER OF WAYS	R	NP	9.3	7.5	0.0	not	0.0	2.2	1.2	not	0.0	7.4	**4.9**	**<.05**	0.0	10.5	**6.2**	**<.05**
A ONE STANDARD DEVIATION INCREASE	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	33.4	**22.0**	**<0.0001**	0.0	0.0	0.0	not
A ONE STANDARD DEVIATION INCREASE IN	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	31.6	**20.8**	**<0.0001**	0.0	0.0	0.0	not
A POSITIVE AND SIGNIFICANT	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	35.3	**23.2**	**<0.0001**	0.0	6.3	3.7	not
A POSITIVE CORRELATION BETWEEN	R	NP	18.5	3.8	2.6	not	0.0	0.0	0.0	not	11.0	5.6	0.9	not	19.8	4.2	**4.9**	**<.05**
A POSITIVE RELATIONSHIP BETWEEN	T	NP	0.0	26.4	**8.3**	**<.01**	0.0	2.2	1.2	not	6.6	22.3	**4.3**	**<.05**	0.0	21.1	**12.3**	**<.001**
A REDUCTION IN THE	R	NP	18.5	3.8	2.6	not	0.0	0.0	0.0	not	13.2	7.4	0.8	not	24.7	4.2	**7.2**	**<.01**
A RESPONSE RATE OF	R	NP	0.0	0.0	0.0	not	0.0	42.3	**23.1**	**<0.0001**	2.2	0.0	1.6	not	0.0	0.0	0.0	not
A RESULT OF THE	T	NP	74.0	22.6	**7.0**	**<.01**	10.6	15.6	0.4	not	30.9	20.4	1.1	not	32.1	35.8	0.1	not
A SHORT PERIOD OF TIME	R	NP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	6.6	3.7	0.4	not	2.5	0.0	1.6	not
A SIGNIFICANT EFFECT ON THE	R	NP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	11.0	3.7	1.9	not	2.5	4.2	0.2	not

Phrase																		
A SIGNIFICANT IMPACT ON THE	R	NP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	7.4	2.1	1.4	not
A SIGNIFICANT INCREASE IN	R	NP	13.9	3.8	1.5	not	2.7	0.0	1.6	not	46.3	7.4	**15.8**	**<0.0001**	17.3	2.1	**6.1**	**<.05**
A SIGNIFICANT INCREASE IN THE	R	NP	9.3	3.8	0.6	not	2.7	0.0	1.6	not	15.4	1.9	**6.2**	**<.05**	7.4	0.0	**4.7**	**<.05**
A SIGNIFICANT PREDICTOR OF	R	NP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	2.2	24.2	**10.2**	**<.01**	2.5	6.3	0.8	not
A SIGNIFICANT ROLE IN	R	NP	13.9	11.3	0.1	not	0.0	0.0	0.0	not	2.2	3.7	0.2	not	7.4	6.3	0.0	not
A SMALL NUMBER OF	R	NP	18.5	18.8	0.0	not	2.7	17.8	**5.0**	**<.05**	4.4	13.0	2.2	not	12.3	14.8	0.1	not
A SUMMARY OF THE	T	NP	9.3	7.5	0.0	not	5.3	4.5	0.0	not	13.2	16.7	0.2	not	2.5	2.1	0.0	not
A WIDE RANGE OF	R	NP	101.8	52.7	3.8	not	16.0	55.7	**9.4**	**<.01**	22.1	18.6	0.1	not	34.6	33.7	0.0	not
A WIDE VARIETY OF	R	NP	50.9	18.8	3.7	not	5.3	8.9	0.4	not	6.6	9.3	0.2	not	24.7	21.1	0.1	not
A WIDER RANGE OF	R	NP	4.6	7.5	0.2	not	0.0	6.7	3.6	not	0.0	14.9	**9.8**	**<.01**	4.9	25.3	**6.4**	**<.05**
ACCORDING TO THE MANUFACTURER	T	PP	0.0	0.0	0.0	not	135.7	2.2	**71.5**	**<0.0001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
ALSO MORE LIKELY TO	P	NP	4.6	3.8	0.0	not	0.0	6.7	3.6	not	0.0	14.9	**9.8**	**<.01**	0.0	10.5	**6.2**	**<.05**
ALSO PLAY A ROLE IN	T	NP_VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	0.0	8.4	**4.9**	**<.05**
AN ANALYSIS OF THE	R	NP	9.3	0.0	3.2	not	16.0	4.5	2.9	not	2.2	13.0	**4.1**	**<.05**	7.4	4.2	0.4	not
AN ESTIMATE OF THE	R	NP	4.6	0.0	1.6	not	16.0	6.7	1.6	not	6.6	5.6	0.0	not	9.9	0.0	**6.2**	**<.05**
AN IMPORTANT ROLE IN	P	NP	97.1	41.4	**5.6**	**<.05**	2.7	4.5	0.2	not	22.1	16.7	0.4	not	76.6	35.8	**6.7**	**<.01**
AN INCREASE IN THE	R	NP	23.1	22.6	0.0	not	2.7	13.4	3.1	not	39.7	27.9	1.0	not	71.6	4.2	**32.6**	**<0.0001**
AN INCREASE IN THE NUMBER OF	R	NP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	2.5	0.0	1.6	not
AN OVERVIEW OF THE	T	NP	13.9	7.5	0.5	not	8.0	20.0	2.2	not	2.2	11.1	3.2	not	0.0	6.3	3.7	not
AN UNDERSTANDING OF THE	T	NP	37.0	7.5	**5.2**	**<.05**	0.0	4.5	2.4	not	0.0	1.9	1.2	not	12.3	8.4	0.3	not
ANALYSIS WAS PERFORMED USING	R	NP_VP	0.0	0.0	0.0	not	55.9	0.0	**33.0**	**<0.0001**	0.0	1.9	1.2	not	0.0	0.0	0.0	not
AND AN INCREASE IN	R	NP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	11.0	5.6	0.9	not	17.3	0.0	**10.9**	**<.001**
AND ARE MORE LIKELY TO	P	VP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	0.0	7.4	**4.9**	**<.05**	0.0	8.4	**4.9**	**<.05**
AND AS A RESULT	T	PP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	2.2	9.3	2.3	not	4.9	16.9	3.0	not
AND AT THE END OF	R	PP	0.0	0.0	0.0	not	2.7	6.7	0.7	not	6.6	1.9	1.4	not	0.0	2.1	1.2	not
AND AT THE SAME TIME	R	PP	0.0	3.8	1.2	not	8.0	0.0	**4.7**	**<.05**	4.4	3.7	0.0	not	4.9	6.3	0.1	not
AND IN THE CASE OF	T	PP	9.3	0.0	3.2	not	0.0	4.5	2.4	not	4.4	1.9	0.5	not	4.9	0.0	3.1	not
AND ON THE OTHER HAND	T	PP	0.0	3.8	1.2	not	2.7	4.5	0.2	not	0.0	3.7	2.4	not	0.0	2.1	1.2	not
AND THE EXTENT TO WHICH	R	PP	0.0	22.6	**7.1**	**<.01**	0.0	8.9	**4.9**	**<.05**	0.0	3.7	2.4	not	0.0	2.1	1.2	not
AND THE LACK OF	R	NP	9.3	7.5	0.0	not	0.0	2.2	1.2	not	4.4	5.6	0.1	not	7.4	12.6	0.6	not
AND THE LIKELIHOOD OF	R	NP	4.6	3.8	0.0	not	0.0	13.4	**7.3**	**<.01**	0.0	3.7	2.4	not	2.5	2.1	0.0	not
AND THE NUMBER OF	R	NP	23.1	7.5	2.0	not	10.6	29.0	3.5	not	26.5	22.3	0.2	not	7.4	4.2	0.4	not
AND THE PRESENCE OF	R	NP	13.9	11.3	0.1	not	8.0	8.9	0.0	not	8.8	11.1	0.1	not	24.7	2.1	**10.0**	**<.01**
AND THE ROLE OF	R	NP	13.9	15.1	0.0	not	2.7	4.5	0.2	not	0.0	7.4	**4.9**	**<.05**	12.3	16.9	0.3	not
AND THE SIZE OF THE	R	NP	0.0	11.3	3.6	not	2.7	0.0	1.6	not	2.2	1.9	0.0	not	7.4	6.3	0.0	not
AND THE USE OF	R	NP	9.3	15.1	0.3	not	5.3	4.5	0.0	not	0.0	11.1	**7.3**	**<.01**	14.8	27.4	1.7	not
AND TO A LESSER EXTENT	R	NP	4.6	0.0	1.6	not	0.0	4.5	2.4	not	8.8	16.7	1.2	not	9.9	4.2	1.0	not
ANY OF THE OTHER	R	NP	0.0	3.8	1.2	not	2.7	0.0	1.6	not	8.8	14.9	0.8	not	7.4	2.1	1.4	not
APPEARS TO BE A	P	VP	13.9	7.5	0.5	not	0.0	2.2	1.2	not	11.0	18.6	1.0	not	29.6	25.3	0.2	not
APPROVED BY THE ETHICS COMMITTEE OF	P	NP_VP	0.0	0.0	0.0	not	16.0	2.2	**4.9**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
ARE ALSO MORE LIKELY	P	VP	4.6	7.5	0.2	not	0.0	4.5	2.4	not	0.0	9.3	**6.1**	**<.05**	0.0	8.4	**4.9**	**<.05**
ARE ALSO MORE LIKELY TO	P	VP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	9.3	**6.1**	**<.05**	0.0	6.3	3.7	not
ARE ASSOCIATED WITH HIGHER LEVELS OF	R	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	11.1	**7.3**	**<.01**	0.0	0.0	0.0	not
ARE BASED ON THE	T	VP	9.3	7.5	0.0	not	5.3	15.6	2.1	not	6.6	9.3	0.2	not	4.9	8.4	0.4	not
ARE CONSISTENT WITH THE	T	VP	4.6	7.5	0.2	not	0.0	6.7	3.6	not	35.3	35.3	0.0	not	22.2	23.2	0.0	not
ARE EXPECTED TO BE	P	VP	18.5	11.3	0.4	not	0.0	2.2	1.2	not	4.4	1.9	0.5	not	0.0	2.1	1.2	not
ARE FOUND TO BE	P	VP	9.3	3.8	0.6	not	2.7	0.0	1.6	not	8.8	3.7	1.1	not	4.9	6.3	0.1	not
ARE GIVEN IN FIG	T	NP_VP	0.0	0.0	0.0	not	16.0	0.0	**9.4**	**<.01**	35.3	0.0	**25.0**	**<0.0001**	2.5	0.0	1.6	not
ARE GIVEN IN TABLE	T	NP_VP	4.6	0.0	1.6	not	39.9	2.2	**17.3**	**<0.0001**	15.4	7.4	1.4	not	7.4	0.0	**4.7**	**<.05**

ARE ILLUSTRATED IN FIG	T	NP_VP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	37.5	11.1	**7.5**	**<.01**	0.0	0.0	0.0	not
ARE IN GOOD AGREEMENT WITH	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	17.6	0.0	**12.5**	**<.001**	7.4	0.0	**4.7**	**<.05**
ARE IN GOOD AGREEMENT WITH THE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	15.4	0.0	**11.0**	**<.001**	2.5	0.0	1.6	not
ARE IN LINE WITH	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	20.4	**8.1**	**<.01**	7.4	27.4	**5.3**	**<.05**
ARE IN LINE WITH THE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	11.1	3.2	not	0.0	8.4	**4.9**	**<.05**
ARE INCLUDED IN THE	R	VP	0.0	0.0	0.0	not	8.0	37.9	**8.5**	**<.01**	2.2	22.3	**9.2**	**<.01**	4.9	2.1	0.5	not
ARE LESS LIKELY TO	P	VP	0.0	18.8	**6.0**	**<.05**	0.0	22.3	**12.2**	**<.001**	0.0	66.9	**44.0**	**<0.0001**	2.5	54.8	**25.1**	**<0.0001**
ARE LESS LIKELY TO BE	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	12.6	**7.4**	**<.01**
ARE LIKELY TO BE	P	VP	9.3	33.9	3.5	not	2.7	33.4	**12.3**	**<.001**	11.0	11.1	0.0	not	22.2	33.7	1.0	not
ARE LIKELY TO BE MORE	P	VP	0.0	7.5	2.4	not	0.0	4.5	2.4	not	0.0	5.6	3.7	not	0.0	2.1	1.2	not
ARE LIKELY TO HAVE	P	VP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	0.0	3.7	2.4	not	14.8	14.8	0.0	not
ARE LISTED IN TABLE	T	NP_VP	0.0	0.0	0.0	not	61.2	4.5	**24.7**	**<0.0001**	52.9	1.9	**30.4**	**<0.0001**	7.4	0.0	**4.7**	**<.05**
ARE MORE LIKELY THAN	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	14.9	**9.8**	**<.01**	2.5	6.3	0.8	not
ARE MORE LIKELY TO	P	VP	4.6	105.4	**26.2**	**<0.0001**	5.3	78.0	**30.2**	**<0.0001**	2.2	206.3	**125.8**	**<0.0001**	14.8	158.1	**59.1**	**<0.0001**
ARE MORE LIKELY TO BE	P	VP	0.0	30.1	**9.5**	**<.01**	2.7	11.1	2.2	not	0.0	27.9	**18.3**	**<0.0001**	2.5	44.3	**19.3**	**<0.0001**
ARE MORE LIKELY TO ENGAGE IN	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	7.4	**4.9**	**<.05**	0.0	4.2	2.5	not
ARE MORE LIKELY TO HAVE	P	VP	0.0	7.5	2.4	not	0.0	8.9	**4.9**	**<.05**	0.0	18.6	**12.2**	**<.001**	0.0	12.6	**7.4**	**<.01**
ARE MORE LIKELY TO MAKE	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	2.1	1.2	not
ARE MORE LIKELY TO RECEIVE	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	13.0	**8.6**	**<.01**	0.0	8.4	**4.9**	**<.05**
ARE MOST LIKELY TO	P	VP	0.0	3.8	1.2	not	2.7	2.2	0.0	not	11.0	13.0	0.1	not	4.9	8.4	0.4	not
ARE MOST LIKELY TO BE	P	VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	11.0	5.6	0.9	not	2.5	2.1	0.0	not
ARE MUCH MORE LIKELY TO	P	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	2.2	13.0	**4.1**	**<.05**	0.0	2.1	1.2	not
ARE NOT INCLUDED IN THE	R	NP_VP	0.0	0.0	0.0	not	5.3	11.1	0.8	not	0.0	5.6	3.7	not	2.5	0.0	1.6	not
ARE PLOTTED IN FIG	T	NP_VP	0.0	0.0	0.0	not	8.0	0.0	**4.7**	**<.05**	50.7	0.0	**36.0**	**<0.0001**	14.8	0.0	**9.3**	**<.01**
ARE PRESENTED IN FIG	T	NP_VP	4.6	0.0	1.6	not	8.0	0.0	**4.7**	**<.05**	77.2	7.4	**33.9**	**<0.0001**	12.3	0.0	**7.8**	**<.01**
ARE PRESENTED IN TABLE	T	NP_VP	0.0	0.0	0.0	not	21.3	42.3	2.9	not	52.9	96.6	**6.3**	**<.05**	2.5	0.0	1.6	not
ARE PROVIDED IN TABLE	T	NP_VP	0.0	0.0	0.0	not	24.0	6.7	**4.3**	**<.05**	11.0	11.1	0.0	not	0.0	0.0	0.0	not
ARE REPORTED IN TABLE	T	NP_VP	0.0	0.0	0.0	not	0.0	13.4	**7.3**	**<.01**	6.6	44.6	**15.2**	**<0.0001**	0.0	0.0	0.0	not
ARE SHOWN IN FIG	T	NP_VP	9.3	0.0	3.2	not	61.2	0.0	**36.2**	**<0.0001**	315.5	5.6	**198.2**	**<0.0001**	54.3	0.0	**34.1**	**<0.0001**
ARE SHOWN IN FIGURE	T	NP_VP	9.3	0.0	3.2	not	8.0	0.0	**4.7**	**<.05**	97.1	5.6	**50.2**	**<0.0001**	7.4	0.0	**4.7**	**<.05**
ARE SHOWN IN TABLE	T	NP_VP	4.6	0.0	1.6	not	34.6	22.3	1.1	not	48.5	59.5	0.5	not	9.9	2.1	2.4	not
ARE SIGNIFICANTLY MORE LIKELY TO	R	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	11.1	**7.3**	**<.01**	0.0	2.1	1.2	not
ARE SIMILAR TO THOSE	T	NP_VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	13.2	13.0	0.0	not	14.8	6.3	1.6	not
ARE SIMILAR TO THOSE OF	T	NP_VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	6.6	1.9	1.4	not	7.4	2.1	1.4	not
ARE SUMMARIZED IN TABLE	T	NP_VP	0.0	3.8	1.2	not	18.6	2.2	**6.2**	**<.05**	30.9	16.7	2.1	not	4.9	0.0	3.1	not
AS A CONSEQUENCE OF	T	PP	23.1	15.1	0.4	not	0.0	4.5	2.4	not	13.2	7.4	0.8	not	14.8	8.4	0.8	not
AS A CONSEQUENCE OF THE	T	PP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	4.4	1.9	0.5	not	0.0	4.2	2.5	not
AS A FORM OF	T	PP	0.0	22.6	**7.1**	**<.01**	0.0	8.9	**4.9**	**<.05**	0.0	3.7	2.4	not	0.0	10.5	**6.2**	**<.05**
AS A FUNCTION OF	R	PP	87.9	11.3	**16.5**	**<0.0001**	82.5	31.2	**10.0**	**<.01**	209.6	39.0	**64.6**	**<0.0001**	106.2	14.8	**34.8**	**<0.0001**
AS A FUNCTION OF TEMPERATURE	R	PP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	11.0	0.0	**7.8**	**<.01**	4.9	0.0	3.1	not
AS A FUNCTION OF THE	R	PP	27.8	0.0	**9.6**	**<.01**	18.6	11.1	0.8	not	57.4	14.9	**13.4**	**<.001**	22.2	4.2	**6.0**	**<.05**
AS A FUNCTION OF TIME	R	PP	4.6	3.8	0.0	not	18.6	2.2	**6.2**	**<.05**	24.3	0.0	**17.2**	**<0.0001**	7.4	2.1	1.4	not
AS A MEANS OF	R	PP	4.6	15.1	1.4	not	0.0	8.9	**4.9**	**<.05**	0.0	9.3	**6.1**	**<.05**	2.5	16.9	**5.1**	**<.05**
AS A MEANS TO	R	PP	9.3	3.8	0.6	not	0.0	2.2	1.2	not	0.0	1.9	1.2	not	4.9	19.0	3.8	not
AS A MEASURE OF	R	PP	0.0	7.5	2.4	not	18.6	11.1	0.8	not	2.2	13.0	**4.1**	**<.05**	7.4	4.2	0.4	not
AS A PROXY FOR	R	PP	0.0	7.5	2.4	not	0.0	20.0	**10.9**	**<.001**	4.4	11.1	1.5	not	2.5	2.1	0.0	not
AS A RESULT OF	T	PP	111.0	67.8	2.5	not	13.3	51.2	**9.6**	**<.01**	64.0	35.3	**4.2**	**<.05**	88.9	78.0	0.3	not
AS A RESULT OF A	T	PP	4.6	7.5	0.2	not	0.0	4.5	2.4	not	6.6	3.7	0.4	not	7.4	2.1	1.4	not
AS A RESULT OF THE	T	PP	69.4	22.6	**6.1**	**<.05**	8.0	13.4	0.6	not	28.7	13.0	3.0	not	19.8	29.5	0.8	not

AS A RESULT OF THEIR	T	PP	4.6	7.5	0.2	not	0.0	4.5	2.4	not	2.2	3.7	0.2	not	0.0	6.3	3.7	not
AS A RESULT OF THIS	T	PP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	0.0	0.0	not	0.0	6.3	3.7	not
AS A RESULT THE	T	PP	9.3	41.4	**5.1**	**<.05**	16.0	8.9	0.8	not	24.3	18.6	0.4	not	19.8	29.5	0.8	not
AS A RESULT WE	T	PP	0.0	22.6	**7.1**	**<.01**	0.0	20.0	**10.9**	**<.001**	2.2	5.6	0.7	not	12.3	8.4	0.3	not
AS AN INDICATOR OF	T	PP	4.6	0.0	1.6	not	8.0	22.3	2.8	not	2.2	1.9	0.0	not	0.0	6.3	3.7	not
AS CAN BE SEEN	P	VP	0.0	0.0	0.0	not	5.3	15.6	2.1	not	57.4	78.0	1.6	not	32.1	8.4	**6.5**	**<.05**
AS CAN BE SEEN FROM THE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	8.8	13.0	0.4	not	0.0	4.2	2.5	not
AS CAN BE SEEN IN FIG	T	NP_VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	26.5	7.4	**5.7**	**<.05**	17.3	2.1	**6.1**	**<.05**
AS CAN BE SEEN IN FIGURE	T	NP_VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	8.8	7.4	0.1	not	0.0	0.0	0.0	not
AS CAN BE SEEN IN TABLE	T	NP_VP	0.0	0.0	0.0	not	2.7	4.5	0.2	not	0.0	29.7	**19.6**	**<0.0001**	0.0	0.0	0.0	not
AS ILLUSTRATED IN FIG	T	NP_VP	0.0	0.0	0.0	not	10.6	4.5	1.1	not	26.5	5.6	**7.4**	**<.01**	29.6	0.0	**18.6**	**<0.0001**
AS IN THE CASE OF	T	PP	0.0	0.0	0.0	not	0.0	8.9	**4.9**	**<.05**	6.6	5.6	0.0	not	9.9	6.3	0.3	not
AS INDICATED BY THE	T	VP	0.0	7.5	2.4	not	2.7	6.7	0.7	not	15.4	33.4	3.3	not	17.3	8.4	1.4	not
AS OPPOSED TO THE	T	PP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	2.2	9.3	2.3	not	17.3	10.5	0.7	not
AS PART OF A	T	PP	13.9	18.8	0.2	not	2.7	17.8	**5.0**	**<.05**	0.0	9.3	**6.1**	**<.05**	4.9	12.6	1.5	not
AS PART OF THE	T	PP	4.6	30.1	**4.9**	**<.05**	13.3	24.5	1.4	not	2.2	14.9	**5.1**	**<.05**	4.9	14.8	2.2	not
AS SHOWN BY THE	T	NP_VP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	13.2	3.7	2.8	not	17.3	8.4	1.4	not
AS SHOWN IN FIG	T	NP_VP	37.0	3.8	**7.7**	**<.01**	138.4	8.9	**57.8**	**<0.0001**	341.9	48.3	**125.4**	**<0.0001**	148.2	8.4	**68.1**	**<0.0001**
AS SHOWN IN FIGURE	T	NP_VP	13.9	3.8	1.5	not	5.3	6.7	0.1	not	90.5	39.0	**10.5**	**<.01**	17.3	4.2	3.8	not
AS SHOWN IN TABLE	T	NP_VP	4.6	0.0	1.6	not	16.0	20.0	0.2	not	33.1	102.2	**18.0**	**<0.0001**	12.3	10.5	0.1	not
AS SHOWN IN THE	T	NP_VP	0.0	0.0	0.0	not	8.0	6.7	0.0	not	19.9	16.7	0.1	not	12.3	10.5	0.1	not
AS THE DEPENDENT VARIABLE	R	NP	4.6	0.0	1.6	not	2.7	29.0	**10.2**	**<.01**	0.0	55.7	**36.7**	**<0.0001**	0.0	8.4	**4.9**	**<.05**
AS THE NUMBER OF	R	PP	0.0	15.1	**4.8**	**<.05**	18.6	80.2	**16.6**	**<0.0001**	4.4	13.0	2.2	not	4.9	12.6	1.5	not
AS THE RATIO OF	R	PP	9.3	3.8	0.6	not	8.0	8.9	0.0	not	8.8	1.9	2.5	not	7.4	2.1	1.4	not
AS THE UNIT OF ANALYSIS	R	PP	0.0	3.8	1.2	not	2.7	13.4	3.1	not	0.0	1.9	1.2	not	2.5	0.0	1.6	not
AS WELL AS A	T	PP	27.8	11.3	1.7	not	16.0	49.0	**7.1**	**<.01**	28.7	27.9	0.0	not	27.2	35.8	0.5	not
AS WELL AS FOR	T	PP	9.3	7.5	0.0	not	2.7	6.7	0.7	not	4.4	16.7	3.7	not	17.3	6.3	2.3	not
AS WELL AS IN	T	PP	9.3	48.9	**6.9**	**<.01**	5.3	29.0	**7.2**	**<.01**	17.6	24.2	0.5	not	27.2	35.8	0.5	not
AS WELL AS IN THE	T	PP	0.0	18.8	**6.0**	**<.05**	2.7	0.0	1.6	not	6.6	1.9	1.4	not	4.9	8.4	0.4	not
AS WELL AS OTHER	T	PP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	0.0	0.0	0.0	not	7.4	14.8	1.1	not
AS WELL AS THE	T	PP	152.6	97.9	2.9	not	69.2	122.5	**6.1**	**<.05**	77.2	102.2	1.7	not	71.6	73.8	0.0	not
AS WELL AS THEIR	T	PP	13.9	18.8	0.2	not	8.0	11.1	0.2	not	4.4	14.9	2.9	not	7.4	12.6	0.6	not
AS WELL AS TO	T	PP	23.1	18.8	0.1	not	10.6	6.7	0.4	not	0.0	9.3	**6.1**	**<.05**	12.3	19.0	0.6	not
ASSOCIATED WITH HIGHER LEVELS	R	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	13.0	**8.6**	**<.01**	0.0	2.1	1.2	not
ASSOCIATED WITH HIGHER LEVELS OF	R	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	13.0	**8.6**	**<.01**	0.0	2.1	1.2	not
AT A FLOW RATE OF	R	PP	0.0	0.0	0.0	not	47.9	0.0	**28.3**	**<0.0001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
AT A RATE OF	R	PP	13.9	3.8	1.5	not	24.0	0.0	**14.1**	**<.001**	8.8	7.4	0.1	not	7.4	0.0	**4.7**	**<.05**
AT HIGH LEVELS OF	R	PP	0.0	18.8	**6.0**	**<.05**	0.0	0.0	0.0	not	4.4	27.9	**9.1**	**<.01**	7.4	10.5	0.2	not
AT LEAST IN PART	P	PP	13.9	7.5	0.5	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	7.4	14.8	1.1	part
AT LEAST ONE OF THE	R	PP	0.0	0.0	0.0	not	5.3	17.8	2.9	not	4.4	11.1	1.5	not	7.4	6.3	0.0	not
AT ROOM TEMPERATURE FOR	R	PP	0.0	0.0	0.0	not	85.2	0.0	**50.3**	**<0.0001**	0.0	0.0	0.0	not	2.5	0.0	1.6	not
AT THE BEGINNING OF	R	PP	4.6	0.0	1.6	not	16.0	49.0	**7.1**	**<.01**	19.9	20.4	0.0	not	17.3	19.0	0.0	not
AT THE BEGINNING OF THE	R	PP	0.0	0.0	0.0	not	8.0	24.5	3.5	not	4.4	11.1	1.5	not	7.4	14.8	1.1	not
AT THE BOTTOM OF	R	PP	0.0	3.8	1.2	not	13.3	4.5	1.9	not	2.2	13.0	**4.1**	**<.05**	7.4	6.3	0.0	not
AT THE BOTTOM OF THE	R	PP	0.0	3.8	1.2	not	8.0	0.0	**4.7**	**<.05**	2.2	7.4	1.4	not	4.9	6.3	0.1	not
AT THE CENTER OF THE	R	PP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	11.0	0.0	**7.8**	**<.01**	14.8	0.0	**9.3**	**<.01**
AT THE CENTRE OF THE	R	PP	0.0	0.0	0.0	not	10.6	0.0	**6.3**	**<.05**	15.4	1.9	**6.2**	**<.05**	9.9	0.0	**6.2**	**<.05**
AT THE CRACK TIP	R	PP	32.4	0.0	**11.2**	**<.001**	8.0	0.0	**4.7**	**<.05**	13.2	0.0	**9.4**	**<.01**	19.8	0.0	**12.4**	**<.001**
AT THE END OF EACH	R	PP	4.6	0.0	1.6	not	5.3	15.6	2.1	not	4.4	1.9	0.5	not	0.0	0.0	0.0	not

AT THE END OF THE	R	PP	18.5	7.5	1.2	not	37.3	49.0	0.7	not	26.5	27.9	0.0	not	12.3	16.9	0.3	not
AT THE EXPENSE OF	T	PP	23.1	33.9	0.5	not	0.0	2.2	1.2	not	4.4	7.4	0.4	not	12.3	25.3	2.0	not
AT THE EXPENSE OF THE	T	PP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	2.5	4.2	0.2	not
AT THE FRONT OF THE	R	PP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	4.4	5.6	0.1	not	7.4	2.1	1.4	not
AT THE INDIVIDUAL LEVEL	T	PP	0.0	11.3	3.6	not	0.0	31.2	**17.0**	**<0.0001**	0.0	20.4	**13.4**	**<.001**	0.0	14.8	**8.6**	**<.01**
AT THE LEVEL OF	T	PP	9.3	30.1	2.7	not	10.6	22.3	1.7	not	13.2	40.9	**7.2**	**<.01**	9.9	37.9	**7.6**	**<.01**
AT THE LEVEL OF THE	T	PP	0.0	3.8	1.2	not	5.3	8.9	0.4	not	2.2	3.7	0.2	not	2.5	12.6	3.2	not
AT THE NATIONAL LEVEL	T	PP	4.6	11.3	0.7	not	0.0	11.1	**6.1**	**<.05**	0.0	3.7	2.4	not	0.0	8.4	**4.9**	**<.05**
AT THE SAME TIME AS	R	PP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	8.8	9.3	0.0	not	4.9	10.5	0.9	not
AT THE SAME TIME IT IS	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	2.5	6.3	0.8	not
AT THE START OF	R	PP	0.0	0.0	0.0	not	8.0	11.1	0.2	not	13.2	13.0	0.0	not	7.4	6.3	0.0	not
AT THE START OF THE	R	PP	0.0	0.0	0.0	not	8.0	4.5	0.4	not	13.2	7.4	0.8	not	4.9	6.3	0.1	not
AT THE TIME OF	R	PP	4.6	11.3	0.7	not	8.0	89.1	**31.6**	**<0.0001**	19.9	29.7	1.0	not	9.9	19.0	1.3	not
AT THE TIME OF DATA COLLECTION	R	PP	0.0	3.8	1.2	not	0.0	13.4	**7.3**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
AT THE TIME OF THE	R	PP	0.0	7.5	2.4	not	5.3	46.8	**15.1**	**<.001**	2.2	11.1	3.2	not	0.0	6.3	3.7	not
AT THE TIME OF THE STUDY	R	PP	0.0	3.8	1.2	not	2.7	8.9	1.4	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
AT THE TIME OF THE SURVEY	T	PP	0.0	0.0	0.0	not	0.0	11.1	**6.1**	**<.05**	0.0	5.6	3.7	not	0.0	2.1	1.2	not
AT THE TOP OF THE	R	PP	4.6	0.0	1.6	not	8.0	2.2	1.4	not	2.2	1.9	0.0	not	4.9	2.1	0.5	not
AT THE UNIVERSITY OF	R	PP	9.3	0.0	3.2	not	34.6	11.1	**5.2**	**<.05**	0.0	0.0	0.0	not	0.0	2.1	1.2	not
ATTENTION HAS BEEN PAID TO	T	NP_VP	0.0	30.1	**9.5**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
BASED ON THE ASSUMPTION THAT	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	3.7	0.2	not	0.0	2.1	1.2	not
BASED ON THE NUMBER OF	T	NP_VP	0.0	0.0	0.0	not	5.3	13.4	1.4	not	0.0	3.7	2.4	not	0.0	0.0	0.0	not
BASED ON THE RESULTS	T	NP_VP	0.0	0.0	0.0	not	0.0	13.4	**7.3**	**<.01**	0.0	9.3	**6.1**	**<.05**	7.4	8.4	0.0	not
BASED ON THE RESULTS OF THE	T	NP_VP	0.0	0.0	0.0	not	0.0	11.1	**6.1**	**<.05**	0.0	3.7	2.4	not	2.5	2.1	0.0	not
BE AFFECTED BY THE	T	VP	4.6	0.0	1.6	not	2.7	8.9	1.4	not	11.0	0.0	**7.8**	**<.01**	4.9	4.2	0.0	not
BE ATTRIBUTED TO THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	28.7	13.0	3.0	not	66.7	25.3	**8.5**	**<.01**
BE CONCLUDED THAT THE	P	VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	17.6	5.6	3.3	not	24.7	4.2	**7.2**	**<.01**
BE CONSIDERED TO BE	P	VP	9.3	0.0	3.2	not	2.7	6.7	0.7	not	11.0	7.4	0.3	not	2.5	12.6	3.2	not
BE DUE TO THE	T	VP	18.5	3.8	2.6	not	0.0	4.5	2.4	not	35.3	26.0	0.7	not	61.7	37.9	2.5	not
BE DUE TO THE FACT THAT	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	9.3	**6.1**	**<.05**	4.9	6.3	0.1	not
BE EXPLAINED BY THE	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	15.4	22.3	0.6	not	29.6	46.4	1.6	not
BE EXPLAINED BY THE FACT THAT	P	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	5.6	0.1	not	2.5	10.5	2.3	not
BE FOUND IN THE	T	VP	13.9	15.1	0.0	not	26.6	11.1	2.7	not	11.0	7.4	0.3	not	7.4	10.5	0.2	not
BE INCLUDED IN THE	R	VP	0.0	7.5	2.4	not	8.0	15.6	1.0	not	0.0	7.4	**4.9**	**<.05**	0.0	8.4	**4.9**	**<.05**
BE INFLUENCED BY THE	P	VP	4.6	15.1	1.4	not	0.0	6.7	3.6	not	0.0	3.7	2.4	not	14.8	6.3	1.6	not
BE LESS LIKELY TO	P	VP	4.6	3.8	0.0	not	0.0	8.9	**4.9**	**<.05**	2.2	20.4	**8.1**	**<.01**	0.0	23.2	**13.6**	**<.001**
BE MORE LIKELY TO	P	VP	0.0	45.2	**14.3**	**<.001**	0.0	15.6	**8.5**	**<.01**	0.0	31.6	**20.8**	**<0.0001**	9.9	37.9	**7.6**	**<.01**
BE NEGATIVELY RELATED TO	R	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	13.0	**8.6**	**<.01**	0.0	6.3	3.7	not
BE NOTED THAT THE	P	VP	9.3	0.0	3.2	not	24.0	6.7	**4.3**	**<.05**	46.3	16.7	**7.2**	**<.01**	32.1	12.6	**3.9**	**<.05**
BE ONE OF THE	P	NP_VP	4.6	11.3	0.7	not	0.0	2.2	1.2	not	2.2	5.6	0.7	not	9.9	10.5	0.0	not
BE POSITIVELY RELATED TO	R	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	14.9	**9.8**	**<.01**	0.0	6.3	3.7	not
BE RELATED TO THE	P	VP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	28.7	11.1	**4.0**	**<.05**	39.5	16.9	**4.1**	**<.05**
BE SEEN IN FIG	T	NP_VP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	55.2	13.0	**14.1**	**<.001**	24.7	2.1	**10.0**	**<.01**
BE SEEN IN THE	T	VP	0.0	0.0	0.0	not	5.3	2.2	0.5	not	6.6	18.6	2.9	not	2.5	6.3	0.8	not
BE TAKEN INTO ACCOUNT	R	VP	4.6	7.5	0.2	not	2.7	2.2	0.0	not	6.6	1.9	1.4	not	7.4	23.2	3.7	not
BE THE RESULT OF	T	NP_VP	4.6	3.8	0.0	not	0.0	8.9	**4.9**	**<.05**	13.2	5.6	1.6	not	9.9	16.9	0.8	not
BE USED AS A	R	VP	4.6	7.5	0.2	not	2.7	8.9	1.4	not	0.0	1.9	1.2	not	14.8	14.8	0.0	not
BEEN FOUND TO BE	T	VP	32.4	22.6	0.4	not	0.0	11.1	**6.1**	**<.05**	0.0	1.9	1.2	not	9.9	23.2	2.4	not
BEEN SHOWN TO BE	T	VP	78.6	15.1	**11.6**	**<.001**	8.0	6.7	0.0	not	8.8	5.6	0.4	not	29.6	10.5	**4.2**	**<.05**

BEFORE AND AFTER THE	R	PP	0.0	11.3	3.6	not	10.6	15.6	0.4	not	11.0	20.4	1.4	not	2.5	4.2	0.2	not
BETWEEN MEN AND WOMEN	T	PP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	2.2	18.6	**7.1**	**<.01**	0.0	2.1	1.2	not
BEYOND THE SCOPE OF THE	T	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	9.9	10.5	0.0	not
BEYOND THE SCOPE OF THIS PAPER	T	PP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	1.9	0.0	not	2.5	10.5	2.3	not
BEYOND THE SCOPE OF THIS STUDY	T	PP	0.0	0.0	0.0	not	0.0	8.9	**4.9**	**<.05**	2.2	0.0	1.6	not	2.5	8.4	1.5	not
BODY OF RESEARCH HAS	T	NP_VP	4.6	33.9	**5.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
BOTH POSITIVE AND NEGATIVE	R	NP	0.0	37.6	**11.9**	**<.001**	0.0	2.2	1.2	not	2.2	1.9	0.0	not	2.5	29.5	**11.5**	**<.001**
BOTH SIDES OF THE	R	NP	0.0	7.5	2.4	not	8.0	6.7	0.0	not	13.2	0.0	**9.4**	**<.01**	4.9	6.3	0.1	not
BUT AT THE SAME TIME	T	PP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	0.0	3.7	2.4	not	0.0	8.4	**4.9**	**<.05**
BY A FACTOR OF	R	PP	4.6	0.0	1.6	not	10.6	0.0	**6.3**	**<.05**	26.5	5.6	**7.4**	**<.01**	22.2	0.0	**14.0**	**<.001**
BY A NUMBER OF	R	PP	13.9	7.5	0.5	not	2.7	2.2	0.0	not	4.4	5.6	0.1	not	4.9	4.2	0.0	not
BY MEANS OF A	R	PP	9.3	3.8	0.6	not	42.6	13.4	**6.7**	**<.01**	4.4	5.6	0.1	not	7.4	0.0	**4.7**	**<.05**
BY THE ADDITION OF	R	PP	4.6	0.0	1.6	not	55.9	0.0	**33.0**	**<0.0001**	6.6	0.0	**4.7**	**<.05**	9.9	0.0	**6.2**	**<.05**
BY THE END OF	R	PP	0.0	3.8	1.2	not	2.7	8.9	1.4	not	6.6	7.4	0.0	not	7.4	6.3	0.0	not
BY THE END OF THE	R	PP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	6.6	1.9	1.4	not	4.9	6.3	0.1	not
BY THE FACT THAT	P	PP	4.6	15.1	1.4	not	0.0	4.5	2.4	not	17.6	26.0	0.8	not	17.3	25.3	0.7	not
BY THE FACT THAT THE	T	PP	4.6	11.3	0.7	not	0.0	0.0	0.0	not	8.8	9.3	0.0	not	2.5	4.2	0.2	not
BY THE NUMBER OF	R	PP	0.0	7.5	2.4	not	8.0	66.8	**21.1**	**<0.0001**	11.0	11.1	0.0	not	2.5	12.6	3.2	not
BY THE PRESENCE OF	R	PP	23.1	0.0	**8.0**	**<.01**	8.0	0.0	**4.7**	**<.05**	26.5	3.7	**9.7**	**<.01**	27.2	6.3	**6.2**	**<.05**
BY THE TOTAL NUMBER OF	R	PP	0.0	0.0	0.0	not	2.7	26.7	**9.1**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
CAN BE APPLIED TO	R	VP	9.3	3.8	0.6	not	0.0	0.0	0.0	not	2.2	3.7	0.2	not	12.3	2.1	3.6	not
CAN BE ATTRIBUTED TO	P	VP	13.9	3.8	1.5	not	2.7	0.0	1.6	not	28.7	9.3	**5.2**	**<.05**	32.1	14.8	2.9	not
CAN BE ATTRIBUTED TO THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	15.4	7.4	1.4	not	24.7	10.5	2.6	not
CAN BE CONSIDERED AS	P	VP	9.3	0.0	3.2	not	8.0	6.7	0.0	not	8.8	3.7	1.1	not	9.9	6.3	0.3	not
CAN BE DEFINED AS	P	VP	9.3	15.1	0.3	not	2.7	6.7	0.7	not	0.0	1.9	1.2	not	0.0	2.1	1.2	not
CAN BE DESCRIBED AS	T	VP	23.1	0.0	**8.0**	**<.01**	2.7	4.5	0.2	not	2.2	1.9	0.0	not	4.9	10.5	0.9	not
CAN BE DIVIDED INTO	P	VP	23.1	0.0	**8.0**	**<.01**	5.3	4.5	0.0	not	8.8	0.0	**6.3**	**<.05**	9.9	4.2	1.0	not
CAN BE EXPLAINED BY	P	VP	4.6	0.0	1.6	not	0.0	4.5	2.4	not	8.8	16.7	1.2	not	37.0	23.2	1.4	not
CAN BE EXPLAINED BY THE	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	4.4	5.6	0.1	not	14.8	21.1	0.5	not
CAN BE EXPRESSED AS	P	VP	4.6	0.0	1.6	not	24.0	4.5	**6.2**	**<.05**	8.8	0.0	**6.3**	**<.05**	17.3	0.0	**10.9**	**<.001**
CAN BE FOUND IN	T	VP	23.1	26.4	0.1	not	63.9	22.3	**8.7**	**<.01**	19.9	18.6	0.0	not	14.8	4.2	2.8	not
CAN BE FOUND IN THE	T	VP	13.9	7.5	0.5	not	26.6	8.9	3.8	not	11.0	5.6	0.9	not	7.4	2.1	1.4	not
CAN BE SEEN AS	P	VP	0.0	22.6	**7.1**	**<.01**	0.0	2.2	1.2	not	2.2	13.0	**4.1**	**<.05**	0.0	19.0	**11.1**	**<.001**
CAN BE SEEN AS A	P	VP	0.0	15.1	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	5.6	3.7	not	0.0	2.1	1.2	not
CAN BE USED AS	R	VP	4.6	11.3	0.7	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	14.8	6.3	1.6	not
CAN BE USED FOR	R	VP	18.5	3.8	2.6	not	8.0	4.5	0.4	not	0.0	0.0	0.0	not	17.3	4.2	3.8	not
CAN BE USED TO	R	VP	83.3	37.6	**4.3**	**<.05**	37.3	17.8	2.9	not	35.3	18.6	2.6	not	42.0	21.1	3.1	not
CAN BE USED TO PREDICT	R	VP	9.3	11.3	0.0	not	2.7	2.2	0.0	not	0.0	1.9	1.2	not	4.9	4.2	0.0	not
CAN BE VIEWED AS	P	VP	0.0	18.8	**6.0**	**<.05**	0.0	4.5	2.4	not	0.0	0.0	0.0	not	4.9	6.3	0.1	not
CONSISTENT WITH THE FINDINGS OF	T	NP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	8.8	0.0	**6.3**	**<.05**	0.0	10.5	**6.2**	**<.05**
CONSISTENT WITH THE RESULTS	T	NP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	13.2	9.3	0.3	not	24.7	6.3	**5.2**	**<.05**
CONSISTENT WITH THE RESULTS OF	T	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	1.9	0.5	not	19.8	4.2	**4.9**	**<.05**
CONTRIBUTE TO A BETTER UNDERSTANDING OF	T	NP_VP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	4.2	0.2	not
CONTRIBUTES TO THE LITERATURE BY	T	NP_VP	0.0	22.6	**7.1**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	8.4	**4.9**	**<.05**
CONTRIBUTION OF THIS STUDY IS	T	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	19.0	**11.1**	**<.001**
COULD BE ARGUED THAT	P	VP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	2.2	11.1	3.2	not	4.9	14.8	2.2	not
COULD BE ATTRIBUTED TO THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	12.3	6.3	0.9	not
COULD BE CONSIDERED AS	P	VP	4.6	0.0	1.6	not	10.6	0.0	**6.3**	**<.05**	24.3	1.9	**11.6**	**<.001**	4.9	2.1	0.5	not
COULD BE DUE TO	P	VP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	22.1	7.4	3.8	not	14.8	8.4	0.8	not

217

Phrase																		
COULD BE DUE TO THE	T	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	4.4	5.6	0.1	not	2.5	6.3	0.8	not
DATA DO NOT ALLOW US TO	R	NP_VP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	6.3	3.7	not
DATA WERE COLLECTED FROM	R	NP_VP	0.0	3.8	1.2	not	8.0	29.0	**5.1**	**<.05**	4.4	0.0	3.1	not	0.0	2.1	1.2	not
DAY TO DAY VARIABILITY	R	NP	13.9	0.0	**4.8**	**<.05**	2.7	0.0	1.6	not	28.7	0.0	**20.3**	**<0.0001**	12.3	0.0	**7.8**	**<.01**
DEGREE TO WHICH THE	R	NP	0.0	7.5	2.4	not	2.7	13.4	3.1	not	0.0	11.1	**7.3**	**<.01**	2.5	14.8	**4.2**	**<.05**
DESPITE THE FACT THAT	T	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	8.8	14.9	0.8	not	14.8	12.6	0.1	not
DID NOT APPEAR TO BE	P	VP	0.0	7.5	2.4	not	2.7	4.5	0.2	not	2.2	1.9	0.0	not	7.4	0.0	**4.7**	**<.05**
DID NOT HAVE A	T	VP	4.6	0.0	1.6	not	5.3	13.4	1.4	not	13.2	22.3	1.1	not	12.3	21.1	1.0	not
DIFFERENCES BETWEEN THE TWO	R	NP	0.0	7.5	2.4	not	2.7	4.5	0.2	not	13.2	18.6	0.4	not	4.9	12.6	1.5	not
DO NOT APPEAR TO BE	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	1.9	1.2	not	9.9	0.0	**6.2**	**<.05**
DOES NOT APPEAR TO	P	VP	9.3	7.5	0.0	not	2.7	2.2	0.0	not	11.0	9.3	0.1	not	12.3	12.6	0.0	not
DUE TO A LACK OF	T	PP	0.0	0.0	0.0	not	2.7	4.5	0.2	not	4.4	1.9	0.5	not	4.2	2.5	0.0	not
DUE TO DIFFERENCES IN	T	NP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	9.3	**6.1**	**<.05**	14.8	2.1	**4.8**	**<.05**
DUE TO THE FACT THAT	T	PP	4.6	7.5	0.2	not	2.7	6.7	0.7	that	11.0	18.6	1.0	not	24.7	19.0	0.3	not
DUE TO THE FACT THAT THE	T	PP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	6.6	7.4	0.0	not	7.4	8.4	0.0	not
DUE TO THE FORMATION OF	T	PP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	6.6	0.0	**4.7**	**<.05**	7.4	0.0	**4.7**	**<.05**
DUE TO THE HIGH	T	NP	32.4	0.0	**11.2**	**<.001**	2.7	2.2	0.0	not	17.6	0.0	**12.5**	**<.001**	27.2	4.2	**8.4**	**<.01**
DUE TO THE LACK OF	T	PP	4.6	3.8	0.0	not	8.0	6.7	0.0	not	8.8	5.6	0.4	not	22.2	4.2	**6.0**	**<.05**
DUE TO THE PRESENCE OF	R	PP	13.9	0.0	**4.8**	**<.05**	8.0	4.5	0.4	not	8.8	1.9	2.5	not	17.3	0.0	**10.9**	**<.001**
DURING THE COURSE OF	R	PP	9.3	0.0	3.2	not	13.3	4.5	1.9	not	15.4	3.7	**3.9**	**<.05**	14.8	0.0	**9.3**	**<.01**
DURING THE COURSE OF THE	R	PP	0.0	0.0	0.0	not	8.0	2.2	1.4	not	8.8	1.9	2.5	not	4.9	0.0	3.1	not
DURING THE PERIOD OF	R	PP	4.6	3.8	0.0	not	13.3	6.7	0.9	not	11.0	1.9	3.6	not	9.9	0.0	**6.2**	**<.05**
DURING THE STUDY PERIOD	R	PP	4.6	0.0	1.6	not	8.0	11.1	0.2	not	13.2	1.9	**4.9**	**<.05**	4.9	8.4	0.4	not
EACH OF THE FOUR	R	NP	0.0	0.0	0.0	not	16.0	22.3	0.4	not	4.4	13.0	2.2	not	0.0	0.0	0.0	not
EACH OF THE THREE	R	NP	0.0	3.8	1.2	not	18.6	24.5	0.3	not	6.6	33.4	**9.5**	**<.01**	4.9	4.2	0.0	not
EVEN IN THE ABSENCE OF	R	NP	13.9	0.0	**4.8**	**<.05**	2.7	4.5	0.2	not	4.4	1.9	0.5	not	4.9	2.1	0.5	not
EXAMINE THE EXTENT TO WHICH	R	NP_VP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	0.0	1.9	1.2	not	0.0	2.1	1.2	not
EXPLAINED BY THE FACT THAT	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	5.6	0.1	not	2.5	10.5	2.3	not
EXTENT TO WHICH A	R	NP	4.6	11.3	0.7	not	0.0	24.5	**13.4**	**<.001**	0.0	11.1	**7.3**	**<.01**	2.5	8.4	1.5	not
EXTENT TO WHICH THE	R	NP	0.0	60.2	**19.1**	**<0.0001**	2.7	55.7	**23.5**	**<0.0001**	0.0	22.3	**14.7**	**<.001**	9.9	14.8	0.4	not
EXTENT TO WHICH THEY	R	NP	0.0	15.1	**4.8**	**<.05**	0.0	22.3	**12.2**	**<.001**	0.0	3.7	2.4	not	0.0	6.3	3.7	not
FINDINGS OF THE PRESENT STUDY	T	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	23.2	**8.2**	**<.01**
FINDINGS OF THIS STUDY	T	NP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	0.0	88.5	**51.8**	**<0.0001**
FINDINGS SUGGEST THAT THE	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	9.3	2.3	not	9.9	16.9	0.8	not
FIRMS ARE MORE LIKELY TO	P	NP_VP	0.0	3.8	1.2	not	0.0	8.9	**4.9**	**<.05**	0.0	3.7	2.4	not	0.0	8.4	**4.9**	**<.05**
FIRST AND THE SECOND	R	NP	18.5	0.0	**6.4**	**<.05**	13.3	0.0	**7.9**	**<.01**	13.2	1.9	**4.9**	**<.05**	9.9	0.0	**6.2**	**<.05**
FIT TO THE DATA	R	NP_VP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	8.8	31.6	**6.6**	**<.05**	0.0	2.1	1.2	not
FOCUSED ON THE ROLE OF	R	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
FOR A NUMBER OF	R	PP	18.5	7.5	1.2	not	2.7	15.6	**4.1**	**<.05**	4.4	9.3	0.9	not	0.0	6.3	3.7	not
FOR A NUMBER OF REASONS	T	PP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	3.7	2.4	not	0.0	2.1	1.2	not
FOR A TOTAL OF	R	PP	0.0	0.0	0.0	not	29.3	17.8	1.2	not	2.2	1.9	0.0	not	0.0	0.0	0.0	not
FOR A VARIETY OF	R	PP	18.5	3.8	2.6	not	2.7	8.9	1.4	not	4.4	0.0	3.1	not	2.5	8.4	1.5	not
FOR A WIDE RANGE OF	R	PP	13.9	0.0	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	7.4	2.1	1.4	not
FOR ALL OF THE	R	PP	0.0	3.8	1.2	not	10.6	8.9	0.1	not	30.9	3.7	**12.3**	**<.001**	9.9	6.3	0.3	not
FOR EACH OF THE	R	PP	0.0	3.8	1.2	not	82.5	106.9	1.3	not	44.1	55.7	0.7	not	9.9	8.4	0.1	not
FOR EACH OF THE THREE	R	PP	0.0	0.0	0.0	not	8.0	11.1	0.2	not	4.4	13.0	2.2	not	2.5	0.0	1.6	not
FOR EACH OF THESE	R	PP	0.0	0.0	0.0	not	10.6	17.8	0.7	not	6.6	3.7	0.4	not	0.0	2.1	1.2	not
FOR EXAMPLE IN THE	T	PP	9.3	11.3	0.0	not	2.7	6.7	0.7	not	2.2	14.9	**5.1**	**<.05**	4.9	8.4	0.4	not
FOR THE ANALYSIS OF	R	PP	9.3	3.8	0.6	not	13.3	6.7	0.9	not	2.2	7.4	1.4	not	0.0	4.2	2.5	not

FOR THE CURRENT STUDY	R	PP	0.0	11.3	3.6	not	5.3	29.0	**7.2**	**<.01**	0.0	1.9	1.2	not	2.5	0.0	1.6	not
FOR THE DEVELOPMENT OF	R	PP	13.9	18.8	0.2	not	2.7	4.5	0.2	not	0.0	0.0	0.0	not	9.9	6.3	0.3	not
FOR THE EFFECT OF	R	PP	0.0	7.5	2.4	not	8.0	20.0	2.2	not	6.6	18.6	2.9	not	4.9	12.6	1.5	not
FOR THE EFFECTS OF	R	PP	0.0	15.1	**4.8**	**<.05**	5.3	17.8	2.9	not	2.2	13.0	**4.1**	**<.05**	2.5	2.1	0.0	not
FOR THE FIRST TIME	R	PP	32.4	18.8	0.9	not	0.0	15.6	**8.5**	**<.01**	4.4	11.1	1.5	not	44.5	16.9	**5.7**	**<.05**
FOR THE FORMATION OF	R	PP	27.8	0.0	**9.6**	**<.01**	0.0	0.0	0.0	not	11.0	0.0	**7.8**	**<.01**	17.3	0.0	**10.9**	**<.001**
FOR THE MAJORITY OF	R	PP	13.9	0.0	**4.8**	**<.05**	5.3	6.7	0.1	not	11.0	3.7	1.9	not	4.9	6.3	0.1	not
FOR THE MOST PART	T	PP	4.6	11.3	0.7	not	0.0	8.9	**4.9**	**<.05**	4.4	11.1	1.5	not	2.5	8.4	1.5	not
FOR THE NUMBER OF	R	PP	0.0	0.0	0.0	not	5.3	31.2	**8.1**	**<.01**	4.4	11.1	1.5	not	0.0	4.2	2.5	not
FOR THE PURPOSE OF THIS STUDY	T	PP	0.0	0.0	0.0	not	0.0	8.9	**4.9**	**<.05**	2.2	1.9	0.0	not	0.0	0.0	0.0	not
FOR THE PURPOSES OF	T	PP	9.3	11.3	0.0	not	18.6	40.1	3.2	not	4.4	9.3	0.9	not	2.5	0.0	1.6	not
FOR THE PURPOSES OF THIS STUDY	T	PP	0.0	0.0	0.0	not	8.0	8.9	0.0	not	0.0	1.9	1.2	not	0.0	0.0	0.0	not
FOR THE RELATIONSHIP BETWEEN	T	PP	0.0	3.8	1.2	not	2.7	11.1	2.2	not	0.0	9.3	**6.1**	**<.05**	2.5	6.3	0.8	not
FOR THE SAKE OF	T	PP	4.6	0.0	1.6	not	5.3	13.4	1.4	not	6.6	14.9	1.6	not	2.5	6.3	0.8	not
FROM A VARIETY OF	R	PP	4.6	15.1	1.4	not	0.0	8.9	**4.9**	**<.05**	0.0	0.0	0.0	not	4.9	10.5	0.9	not
FROM THAT OF THE	R	PP	4.6	0.0	1.6	not	5.3	4.5	0.0	not	24.3	5.6	**6.3**	**<.05**	12.3	0.0	**7.8**	**<.01**
FROM THE PERSPECTIVE OF	T	PP	9.3	15.1	0.3	not	2.7	8.9	1.4	not	0.0	9.3	**6.1**	**<.05**	2.5	21.1	**7.2**	**<.01**
FROM THE PERSPECTIVE OF THE	T	PP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	1.9	1.2	not	0.0	10.5	**6.2**	**<.05**
FROM THE POINT OF VIEW OF	T	PP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	0.0	1.9	1.2	not	7.4	0.0	**4.7**	**<.05**
FURTHER RESEARCH IS NEEDED TO	P	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	4.9	14.8	2.2	not
GAIN A BETTER UNDERSTANDING OF	T	NP_VP	4.6	3.8	0.0	not	0.0	2.2	1.2	not	2.2	0.0	1.6	not	0.0	6.3	3.7	not
GOAL OF THIS STUDY IS TO	T	NP_VP	13.9	3.8	1.5	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
GOOD AGREEMENT WITH THE	T	NP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	44.1	0.0	**31.3**	**<0.0001**	22.2	0.0	**14.0**	**<.001**
HAD A SIGNIFICANT EFFECT ON	R	NP_VP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	11.0	3.7	1.9	not	0.0	2.1	1.2	not
HAS A POSITIVE EFFECT ON	R	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	1.9	1.2	not	0.0	2.1	1.2	not
HAS ALSO BEEN SHOWN TO	T	VP	18.5	0.0	**6.4**	**<.05**	2.7	0.0	1.6	not	0.0	0.0	0.0	not	12.3	2.1	3.6	not
HAS BEEN FOUND TO	T	VP	32.4	11.3	2.6	not	0.0	13.4	**7.3**	**<.01**	0.0	0.0	0.0	not	14.8	12.6	0.1	not
HAS BEEN FOUND TO BE	T	VP	18.5	7.5	1.2	not	0.0	8.9	**4.9**	**<.05**	0.0	0.0	0.0	not	9.9	12.6	0.1	not
HAS BEEN REPORTED TO	P	VP	27.8	3.8	**5.1**	**<.05**	2.7	0.0	1.6	not	0.0	0.0	0.0	not	29.6	0.0	**18.6**	**<0.0001**
HAS BEEN SHOWN THAT	T	VP	60.1	3.8	**14.8**	**<.001**	5.3	2.2	0.5	not	24.3	3.7	**8.5**	**<.01**	46.9	6.3	**15.6**	**<0.0001**
HAS BEEN SHOWN TO	T	VP	120.3	15.1	**22.9**	**<0.0001**	13.3	24.5	1.4	not	15.4	5.6	2.4	not	66.7	4.2	**29.8**	**<0.0001**
HAS BEEN SHOWN TO BE	T	VP	37.0	7.5	**5.2**	**<.05**	2.7	4.5	0.2	not	4.4	5.6	0.1	not	22.2	0.0	**14.0**	**<.001**
HAS BEEN USED IN	R	VP	0.0	15.1	**4.8**	**<.05**	5.3	33.4	**9.1**	**<.01**	4.4	0.0	3.1	not	2.5	2.1	0.0	not
HAS BEEN USED TO	R	VP	37.0	3.8	**7.7**	**<.01**	5.3	4.5	0.0	not	6.6	1.9	1.4	not	7.4	2.1	1.4	not
HAS IMPORTANT IMPLICATIONS FOR	P	NP_VP	4.6	3.8	0.0	not	0.0	0.0	0.0	not	4.4	0.0	3.1	not	7.4	27.4	**5.3**	**<.05**
HAS TENDED TO FOCUS ON	T	VP	4.6	18.8	2.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	6.3	3.7	not
HAS THE POTENTIAL TO	T	NP_VP	13.9	18.8	0.2	not	10.6	0.0	**6.3**	**<.05**	2.2	1.9	0.0	not	12.3	10.5	0.1	not
HAS YET TO BE	P	VP	4.6	33.9	**5.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.9	4.2	0.0	not
HAVE A NEGATIVE EFFECT ON	R	NP_VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	0.0	0.0	not
HAVE A POSITIVE IMPACT ON	R	NP_VP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	6.3	3.7	not
HAVE A POSITIVE RELATIONSHIP WITH	T	NP_VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	0.0	0.0	not
HAVE A SIGNIFICANT IMPACT ON	R	NP_VP	13.9	0.0	**4.8**	**<.05**	0.0	2.2	1.2	not	2.2	5.6	0.7	not	0.0	0.0	0.0	not
HAVE AN EFFECT ON THE	T	NP_VP	0.0	11.3	3.6	not	2.7	4.5	0.2	not	6.6	0.0	**4.7**	**<.05**	4.9	0.0	3.1	not
HAVE AN IMPACT ON THE	T	NP_VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	4.4	5.6	0.1	not	2.5	0.0	1.6	not
HAVE BEEN FOUND TO	T	VP	32.4	22.6	0.4	not	0.0	6.7	3.6	not	2.2	3.7	0.2	not	2.5	14.8	**4.2**	**<.05**
HAVE BEEN FOUND TO BE	T	VP	9.3	7.5	0.2	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	10.5	**6.2**	**<.05**
HAVE BEEN SHOWN TO	T	VP	78.6	30.1	**5.4**	**<.05**	8.0	11.1	0.2	not	11.0	0.0	**7.8**	**<.01**	39.5	14.8	**5.2**	**<.05**
HAVE BEEN SHOWN TO BE	T	VP	37.0	7.5	**5.2**	**<.05**	2.7	2.2	0.0	not	2.2	0.0	1.6	not	4.9	6.3	0.1	not
HAVE BEEN USED IN	R	VP	23.1	0.0	**8.0**	**<.01**	8.0	13.4	0.6	not	6.6	3.7	0.4	not	2.5	2.1	0.0	not

Phrase																		
HAVE BEEN USED TO	R	VP	27.8	11.3	1.7	not	10.6	4.5	1.1	not	0.0	1.9	1.2	not	17.3	4.2	3.8	not
HAVE FOCUSED ON THE	T	VP	18.5	33.9	1.1	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	21.1	**12.3**	**<.001**
HAVE IMPORTANT IMPLICATIONS FOR	P	NP_VP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	29.5	**11.5**	**<.001**
HAVE SHOWN THAT THE	T	VP	55.5	11.3	**7.8**	**<.01**	0.0	0.0	0.0	not	8.8	0.0	**6.3**	**<.05**	29.6	10.5	**4.2**	**<.05**
HAVE THE POTENTIAL TO	T	NP_VP	18.5	15.1	0.1	not	0.0	4.5	2.4	not	0.0	3.7	2.4	not	9.9	16.9	0.8	not
HIGHER THAN THAT OF	R	PP	37.0	0.0	**12.8**	**<.001**	5.3	2.2	0.5	not	39.7	14.9	**5.9**	**<.05**	46.9	4.2	**18.7**	**<0.0001**
HIGHER THAN THAT OF THE	R	PP	13.9	0.0	**4.8**	**<.05**	5.3	2.2	0.5	not	11.0	7.4	0.3	not	17.3	2.1	**6.1**	**<.05**
HIGHER THAN THOSE IN	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	28.7	1.9	**14.4**	**<.001**	29.6	0.0	**18.6**	**<0.0001**
HIGHER THAN THOSE IN THE	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	17.6	1.9	**7.5**	**<.01**	14.8	0.0	**9.3**	**<.01**
HIGHER THAN THOSE OF	R	PP	13.9	0.0	**4.8**	**<.05**	0.0	0.0	0.0	not	22.1	7.4	3.8	not	17.3	0.0	**10.9**	**<.001**
HIGHLIGHT THE IMPORTANCE OF	P	NP_VP	0.0	18.8	**6.0**	**<.05**	0.0	4.5	2.4	not	0.0	3.7	2.4	not	9.9	16.9	0.8	not
HIGHLIGHTS THE IMPORTANCE OF	P	NP_VP	4.6	11.3	0.7	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	2.5	40.0	**17.1**	**<0.0001**
HOWEVER IT SHOULD BE	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	13.2	3.7	2.8	not	19.8	6.3	3.2	not
HOWEVER IT SHOULD BE NOTED THAT	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	13.2	1.9	**4.9**	**<.05**	9.9	4.2	1.0	not
HOWEVER IT SHOULD BE NOTED THAT THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	8.8	1.9	2.5	not	4.9	2.1	0.5	not
HOWEVER THERE IS A	T	VP	13.9	30.1	1.4	not	0.0	4.5	2.4	not	6.6	7.4	0.0	not	9.9	2.1	2.4	not
IF THIS IS THE CASE	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	3.7	0.2	not	0.0	8.4	**4.9**	**<.05**
IMPORTANT ROLE IN THE	P	NP	32.4	11.3	2.6	not	0.0	6.7	3.6	not	4.4	5.6	0.1	not	37.0	14.8	**4.4**	**<.05**
IMPORTANT TO NOTE THAT	P	VP	4.6	11.3	0.7	not	2.7	15.6	**4.1**	**<.05**	6.6	27.9	**6.8**	**<.01**	19.8	27.4	0.5	not
IN A NUMBER OF	R	PP	41.6	22.6	1.4	not	0.0	13.4	**7.3**	**<.01**	6.6	14.9	1.6	not	17.3	21.1	0.2	not
IN A NUMBER OF WAYS	T	PP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	6.3	3.7	not
IN A PREVIOUS STUDY	T	PP	4.6	0.0	1.6	not	29.3	2.2	**11.6**	**<.001**	11.0	0.0	**7.8**	**<.01**	34.6	0.0	**21.7**	**<0.0001**
IN A RANGE OF	R	PP	18.5	7.5	1.2	not	8.0	6.7	0.0	not	6.6	1.9	1.4	not	9.9	6.3	0.3	not
IN A SIMILAR MANNER	T	PP	13.9	0.0	**4.8**	**<.05**	5.3	4.5	0.0	not	8.8	1.9	2.5	not	12.3	2.1	3.6	not
IN A STUDY OF	T	PP	9.3	3.8	0.6	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	7.4	2.1	1.4	not
IN A VARIETY OF	R	PP	27.8	41.4	0.6	not	0.0	15.6	**8.5**	**<.01**	0.0	5.6	3.7	not	19.8	21.1	0.0	not
IN A VARIETY OF WAYS	R	PP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	2.5	4.2	0.2	not
IN A WAY THAT	T	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	2.2	20.4	**8.1**	**<.01**	2.5	16.9	**5.1**	**<.05**
IN A WAY THAT IS	T	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	0.0	1.6	not	0.0	2.1	1.2	not
IN A WIDE VARIETY OF	R	PP	27.8	0.0	**9.6**	**<.01**	0.0	2.2	1.2	not	0.0	0.0	0.0	not	2.5	4.2	0.2	not
IN ACCORDANCE WITH THE	T	PP	9.3	7.5	0.0	not	37.3	13.4	**4.9**	**<.05**	8.8	13.0	0.4	not	27.2	16.9	1.1	not
IN ADDITION TO THE	T	PP	50.9	22.6	2.7	not	24.0	62.4	**7.1**	**<.01**	28.7	46.5	2.1	not	44.5	33.7	0.6	not
IN AGREEMENT WITH THE	T	PP	0.0	3.8	1.2	not	2.7	0.0	1.6	not	33.1	0.0	**23.5**	**<0.0001**	19.8	0.0	**12.4**	**<.001**
IN ALL OF THE	R	PP	4.6	0.0	1.6	not	13.3	15.6	0.1	not	28.7	11.1	**4.0**	**<.05**	9.9	2.1	2.4	not
IN AN ATTEMPT TO	T	PP	9.3	15.1	0.3	not	0.0	13.4	**7.3**	**<.01**	11.0	9.3	0.1	not	4.9	10.5	0.9	not
IN AN EFFORT TO	R	PP	13.9	30.1	1.4	not	8.0	4.5	0.4	not	6.6	13.0	1.0	not	4.9	14.8	2.2	not
IN ANY OF THE	R	PP	0.0	3.8	1.2	not	2.7	6.7	0.7	not	22.1	22.3	0.0	not	9.9	0.0	**6.2**	**<.05**
IN CONTRAST TO THE	T	PP	9.3	18.8	0.8	not	0.0	2.2	1.2	not	35.3	33.4	0.0	not	39.5	16.9	**4.1**	**<.05**
IN EACH OF THE	R	PP	4.6	3.8	0.0	not	16.0	42.3	**5.0**	**<.05**	26.5	33.4	0.4	not	4.9	4.2	0.0	not
IN FRONT OF THE	R	PP	0.0	0.0	0.0	not	10.6	6.7	0.4	not	24.3	1.9	**11.6**	**<.001**	9.9	2.1	2.4	not
IN GOOD AGREEMENT WITH	T	PP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	48.5	0.0	**34.4**	**<0.0001**	37.0	0.0	**23.3**	**<0.0001**
IN GOOD AGREEMENT WITH THE	T	PP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	37.5	0.0	**26.6**	**<0.0001**	14.8	0.0	**9.3**	**<.01**
IN LIGHT OF THE	T	PP	4.6	33.9	**5.8**	**<.05**	0.0	2.2	1.2	not	6.6	11.1	0.6	not	17.3	25.3	0.7	not
IN LINE WITH PREVIOUS	T	PP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	0.0	16.7	**11.0**	**<.001**	7.4	19.0	2.3	not
IN LINE WITH PREVIOUS STUDIES	P	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	7.4	**4.9**	**<.05**	4.9	6.3	0.1	not
IN LINE WITH THE	T	PP	13.9	11.3	0.1	not	2.7	20.0	**6.0**	**<.05**	8.8	50.2	**15.4**	**<0.0001**	14.8	48.5	**8.1**	**<.01**
IN LINE WITH THIS	T	PP	0.0	15.1	**4.8**	**<.05**	0.0	2.2	1.2	not	0.0	1.9	1.2	not	9.9	6.3	0.3	not
IN ONE OF THE	R	PP	23.1	7.5	2.0	not	5.3	17.8	2.9	not	4.4	9.3	0.9	not	2.5	4.2	0.2	not
IN ORDER TO AVOID	R	NP_VP	0.0	7.5	2.4	not	18.6	24.5	0.3	not	2.2	5.6	0.7	not	4.9	4.2	0.0	not

Phrase																		
IN ORDER TO DETERMINE	R	NP_VP	18.5	3.8	2.6	not	16.0	6.7	1.6	not	15.4	3.7	**3.9**	**<.05**	4.9	0.0	3.1	not
IN ORDER TO DETERMINE THE	R	NP_VP	13.9	0.0	**4.8**	**<.05**	10.6	4.5	1.1	not	2.2	1.9	0.0	not	4.9	0.0	3.1	not
IN ORDER TO EXAMINE	R	NP_VP	0.0	3.8	1.2	not	0.0	8.9	**4.9**	**<.05**	11.0	14.9	0.3	not	0.0	2.1	1.2	not
IN ORDER TO IDENTIFY	R	NP_VP	4.6	7.5	0.2	not	2.7	17.8	**5.0**	**<.05**	2.2	1.9	0.0	not	2.5	2.1	0.0	not
IN ORDER TO INVESTIGATE THE	R	NP_VP	4.6	0.0	1.6	not	16.0	0.0	**9.4**	**<.01**	8.8	1.9	2.5	not	0.0	2.1	1.2	not
IN ORDER TO TEST	R	NP_VP	0.0	11.3	3.6	not	2.7	17.8	**5.0**	**<.05**	2.2	11.1	3.2	not	0.0	0.0	0.0	not
IN ORDER TO UNDERSTAND THE	T	NP_VP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	1.9	1.2	not	4.9	4.2	0.0	not
IN OTHER WORDS THE	T	PP	0.0	22.6	**7.1**	**<.01**	2.7	26.7	**9.1**	**<.01**	4.4	46.5	**19.4**	**<0.0001**	24.7	29.5	0.2	not
IN RELATION TO THE	T	PP	13.9	37.6	2.7	not	10.6	33.4	**5.0**	**<.05**	6.6	33.4	**9.5**	**<.01**	22.2	33.7	1.0	not
IN RESPONSE TO THE	T	PP	13.9	33.9	2.0	not	2.7	17.8	**5.0**	**<.05**	4.4	22.3	**6.3**	**<.05**	19.8	19.0	0.0	not
IN SUCH A WAY	T	PP	9.3	3.8	0.6	not	16.0	0.0	**9.4**	**<.01**	8.8	7.4	0.1	not	17.3	10.5	0.7	not
IN SUCH A WAY THAT	T	PP	9.3	3.8	0.6	not	10.6	0.0	**6.3**	**<.05**	6.6	7.4	0.0	not	12.3	4.2	1.8	not
IN TERMS OF A	T	PP	4.6	11.3	0.7	not	5.3	24.5	**5.4**	**<.05**	6.6	22.3	**4.3**	**<.05**	9.9	29.5	**4.4**	**<.05**
IN TERMS OF ITS	T	PP	4.6	7.5	0.2	not	2.7	8.9	1.4	not	2.2	5.6	0.7	not	0.0	12.6	**7.4**	**<.01**
IN TERMS OF THE	T	PP	41.6	82.8	3.3	not	21.3	75.7	**13.0**	**<.001**	41.9	94.8	**10.2**	**<.01**	22.2	105.4	**25.3**	**<0.0001**
IN TERMS OF THEIR	T	PP	4.6	33.9	**5.8**	**<.05**	8.0	26.7	**4.3**	**<.05**	6.6	20.4	3.6	not	2.5	40.0	**17.1**	**<0.0001**
IN THE ABSENCE OF	R	PP	64.8	15.1	**8.1**	**<.01**	24.0	13.4	1.3	not	161.0	26.0	**54.6**	**<0.0001**	106.2	29.5	**20.4**	**<0.0001**
IN THE ABSENCE OF A	R	PP	27.8	3.8	**5.1**	**<.05**	5.3	4.5	0.0	not	33.1	5.6	**10.9**	**<.001**	24.7	6.3	**5.2**	**<.05**
IN THE ABSENCE OF ANY	R	PP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	8.8	0.0	**6.3**	**<.05**	4.9	2.1	0.5	not
IN THE ABSENCE OF THE	R	PP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	13.2	1.9	**4.9**	**<.05**	7.4	2.1	1.4	not
IN THE AMOUNT OF	R	PP	4.6	3.8	0.0	not	10.6	4.5	1.1	not	17.6	11.1	0.7	not	7.4	10.5	0.2	not
IN THE CASE OF	T	PP	41.6	26.4	0.8	not	31.9	60.1	3.6	not	119.1	98.5	1.0	not	140.8	97.0	3.6	not
IN THE CASE OF A	T	PP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	2.2	7.4	1.4	not	17.3	8.4	1.4	not
IN THE CASE OF THE	T	PP	4.6	7.5	0.2	not	16.0	4.5	2.9	not	35.3	5.6	**12.1**	**<.001**	34.6	21.1	1.5	not
IN THE CENTER OF THE	R	PP	0.0	0.0	0.0	not	13.3	0.0	**7.9**	**<.01**	6.6	0.0	**4.7**	**<.05**	7.4	0.0	**4.7**	**<.05**
IN THE CENTRE OF THE	R	PP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	6.6	3.7	0.4	not	9.9	0.0	**6.2**	**<.05**
IN THE CONTEXT OF	T	PP	64.8	131.8	**5.5**	**<.05**	2.7	49.0	**20.1**	**<0.0001**	13.2	37.2	**5.7**	**<.05**	46.9	109.6	**11.2**	**<.001**
IN THE CONTEXT OF A	T	PP	4.6	18.8	2.2	not	0.0	11.1	**6.1**	**<.05**	2.2	3.7	0.2	not	0.0	12.6	**7.4**	**<.01**
IN THE CONTEXT OF THE	T	PP	4.6	22.6	3.0	not	0.0	8.9	**4.9**	**<.05**	4.4	5.6	0.1	not	14.8	29.5	2.1	not
IN THE CONTROL GROUP	R	PP	0.0	3.8	1.2	not	13.3	13.4	0.0	not	30.9	9.3	**6.1**	**<.05**	27.2	0.0	**17.1**	**<0.0001**
IN THE COURSE OF	R	PP	0.0	3.8	1.2	not	2.7	2.2	0.0	not	6.6	3.7	0.4	not	9.9	8.4	0.1	not
IN THE CURRENT STUDY	T	PP	37.0	18.8	1.5	not	18.6	35.6	2.2	not	11.0	7.4	0.3	not	66.7	42.2	2.5	not
IN THE CURRENT STUDY THE	R	PP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	2.2	0.0	1.6	not	7.4	0.0	**4.7**	**<.05**
IN THE CURRENT STUDY WE	R	PP	9.3	15.1	0.3	not	2.7	4.5	0.2	not	0.0	0.0	0.0	not	4.9	4.2	0.0	not
IN THE DEVELOPMENT OF	R	PP	32.4	22.6	0.4	not	0.0	4.5	2.4	not	4.4	1.9	0.5	not	29.6	16.9	1.6	not
IN THE DIRECTION OF	R	PP	9.3	0.0	3.2	not	0.0	6.7	3.6	not	15.4	1.9	**6.2**	**<.05**	7.4	6.3	0.0	not
IN THE DISTRIBUTION OF	R	PP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	6.6	7.4	0.0	not	2.5	8.4	1.5	not
IN THE EARLY STAGES OF	R	PP	0.0	0.0	0.0	not	5.3	2.2	0.5	not	4.4	1.9	0.5	not	7.4	8.4	0.0	not
IN THE EXPECTED DIRECTION	R	PP	4.6	0.0	1.6	not	0.0	6.7	3.6	not	0.0	44.6	**29.3**	**<0.0001**	0.0	4.2	2.5	not
IN THE EXTENT TO WHICH	R	PP	0.0	15.1	**4.8**	**<.05**	0.0	4.5	2.4	not	0.0	5.6	3.7	not	0.0	8.4	**4.9**	**<.05**
IN THE FACE OF	T	PP	0.0	15.1	**4.8**	**<.05**	0.0	2.2	1.2	not	0.0	11.1	**7.3**	**<.01**	2.5	25.3	**9.3**	**<.01**
IN THE FIELD OF	T	PP	18.5	112.9	**17.5**	**<0.0001**	0.0	20.0	**10.9**	**<.001**	2.2	9.3	2.3	not	2.5	21.1	**7.2**	**<.01**
IN THE FIRST PLACE	R	PP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	7.4	**4.9**	**<.05**	4.9	10.5	0.9	not
IN THE FOLLOWING SECTION	T	PP	0.0	15.1	**4.8**	**<.05**	18.6	13.4	0.4	not	15.4	14.9	0.0	not	4.9	2.1	0.5	not
IN THE FORM OF	R	PP	41.6	37.6	0.0	not	24.0	31.2	0.4	not	19.9	55.7	**8.6**	**<.01**	22.2	52.7	**5.5**	**<.05**
IN THE FORM OF A	R	PP	9.3	7.5	0.2	not	2.7	8.9	1.4	not	2.2	14.9	**5.1**	**<.05**	2.5	4.2	0.2	not
IN THE FORMATION OF	R	PP	23.1	0.0	**8.0**	**<.01**	0.0	4.5	2.4	not	22.1	1.9	**10.2**	**<.01**	34.6	6.3	**9.6**	**<.01**
IN THE LEVEL OF	R	PP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	11.0	16.7	0.6	not	0.0	10.5	**6.2**	**<.05**
IN THE LITERATURE ON	T	PP	0.0	26.4	**8.3**	**<.01**	0.0	13.4	**7.3**	**<.01**	2.2	3.7	0.2	not	2.5	12.6	3.2	not

Phrase																		
IN THE LONG RUN	R	PP	0.0	11.3	3.6	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	21.1	**12.3**	**<.001**
IN THE LONG TERM	R	PP	9.3	3.8	0.6	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	2.5	14.8	**4.2**	**<.05**
IN THE MIDDLE OF THE	R	PP	0.0	0.0	0.0	not	10.6	6.7	0.4	not	15.4	5.6	2.4	not	12.3	0.0	**7.8**	**<.01**
IN THE NEXT SECTION	T	PP	9.3	37.6	**4.3**	**<.05**	8.0	29.0	**5.1**	**<.05**	6.6	22.3	**4.3**	**<.05**	4.9	2.1	0.5	not
IN THE NEXT SECTION WE	T	PP	4.6	30.1	**4.9**	**<.05**	2.7	4.5	0.2	not	4.4	3.7	0.0	not	0.0	0.0	0.0	not
IN THE NORTHERN HEMISPHERE	R	PP	23.1	0.0	**8.0**	**<.01**	8.0	0.0	**4.7**	**<.05**	37.5	0.0	**26.6**	**<0.0001**	19.8	0.0	**12.4**	**<.001**
IN THE NUMBER OF	R	PP	0.0	11.3	3.6	not	8.0	8.9	0.0	not	41.9	26.0	1.9	not	34.6	12.6	**4.7**	**<.05**
IN THE OPPOSITE DIRECTION	R	PP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	11.0	18.6	1.0	not	4.9	6.3	0.1	not
IN THE PRESENCE AND ABSENCE OF	R	PP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	17.6	0.0	**12.5**	**<.001**	4.9	0.0	3.1	not
IN THE PRESENCE OF	R	PP	83.3	41.4	3.5	not	79.8	13.4	**22.0**	**<0.0001**	236.1	31.6	**89.1**	**<0.0001**	93.8	31.6	**14.3**	**<.001**
IN THE PRESENCE OF A	R	PP	9.3	18.8	0.8	not	26.6	2.2	**10.2**	**<.01**	26.5	5.6	**7.4**	**<.01**	4.9	4.2	0.0	not
IN THE PRESENCE OF HIGH	R	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	4.4	9.3	**4.9**	**<.05**	0.0	8.4	**4.9**	**<.05**
IN THE PRESENCE OF THE	R	PP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	22.1	3.7	**7.3**	**<.01**	22.2	4.2	**6.0**	**<.05**
IN THE PRESENCE OR ABSENCE OF	R	PP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	11.0	1.9	3.6	not	2.5	2.1	0.0	not
IN THE PRESENT STUDY	T	PP	78.6	48.9	1.7	not	69.2	62.4	0.1	not	28.7	20.4	0.7	not	244.5	109.6	**23.3**	**<0.0001**
IN THE PRESENT STUDY WE	T	PP	27.8	30.1	0.0	not	0.0	6.7	3.6	not	4.4	0.0	3.1	not	19.8	8.4	2.1	not
IN THE PRESENT WORK	R	PP	32.4	0.0	**11.2**	**<.001**	16.0	2.2	**4.9**	**<.05**	15.4	0.0	**11.0**	**<.001**	56.8	0.0	**35.7**	**<0.0001**
IN THE PREVIOUS SECTION	T	PP	0.0	0.0	0.0	not	10.6	6.7	0.4	not	26.5	9.3	**4.3**	**<.05**	9.9	6.3	0.3	not
IN THE PROBABILITY OF	R	PP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	4.4	48.3	**20.5**	**<0.0001**	0.0	4.2	2.5	not
IN THE PROCESS OF	R	PP	9.3	26.4	2.0	not	0.0	4.5	2.4	not	4.4	9.3	0.9	not	12.3	16.9	0.3	not
IN THE RANGE OF	R	PP	41.6	0.0	**14.4**	**<.001**	31.9	2.2	**13.0**	**<.001**	59.6	11.1	**18.3**	**<0.0001**	76.6	0.0	**48.1**	**<0.0001**
IN THE SAME MANNER	T	PP	4.6	3.8	0.0	not	18.6	8.9	1.4	not	2.2	3.7	0.2	not	7.4	0.0	**4.7**	**<.05**
IN THE SAME WAY	T	PP	0.0	18.8	**6.0**	**<.05**	10.6	15.6	0.4	not	4.4	18.6	**4.5**	**<.05**	9.9	10.5	0.0	not
IN THE SAME WAY AS	T	PP	0.0	3.8	1.2	not	2.7	2.2	0.0	not	0.0	9.3	**6.1**	**<.05**	4.9	4.2	0.0	not
IN THE SENSE THAT	T	PP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	6.6	14.9	1.6	not	2.5	29.5	**11.5**	**<.001**
IN THE SENSE THAT IT	T	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	2.5	10.5	2.3	not
IN THE STUDY OF	T	PP	13.9	37.6	2.7	not	2.7	8.9	1.4	not	4.4	0.0	3.1	not	22.2	19.0	0.1	not
IN THE SUPPORTING INFORMATION	T	PP	0.0	0.0	0.0	not	39.9	6.7	**11.0**	**<.001**	39.7	5.6	**14.6**	**<.001**	2.5	2.1	0.0	not
IN THE USE OF	R	PP	0.0	26.4	**8.3**	**<.01**	0.0	22.3	**12.2**	**<.001**	4.4	29.7	**10.1**	**<.01**	2.5	19.0	**6.2**	**<.05**
IN THE VICINITY OF	R	PP	13.9	0.0	**4.8**	**<.05**	5.3	0.0	3.1	not	33.1	0.0	**23.5**	**<0.0001**	29.6	0.0	**18.6**	**<0.0001**
IN THE VICINITY OF THE	R	PP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	11.0	0.0	**7.8**	**<.01**	17.3	0.0	**10.9**	**<.001**
IN THEIR STUDY OF	T	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	4.9	6.3	0.1	not
IN THIS CASE THE	T	PP	0.0	11.3	3.6	not	8.0	13.4	0.6	not	30.9	18.6	1.5	not	22.2	12.6	1.2	not
IN THIS SECTION WE	T	PP	0.0	3.8	1.2	not	5.3	4.5	0.0	not	26.5	27.9	0.0	not	0.0	4.2	2.5	not
IN THIS STUDY TO	R	PP	4.6	0.0	1.6	not	16.0	6.7	1.6	not	2.2	1.9	0.0	not	12.3	8.4	0.3	not
IN THIS STUDY WE	T	PP	134.1	64.0	**6.1**	**<.05**	58.6	37.9	1.8	not	17.6	0.0	**12.5**	**<.001**	79.0	63.2	0.8	not
IN THIS STUDY WE EXAMINE THE	T	NP_VP	4.6	15.1	1.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
INCLUDED IN THE ANALYSIS	R	NP_VP	0.0	7.5	2.4	not	24.0	26.7	0.1	not	0.0	5.6	3.7	not	2.5	4.2	0.2	not
INCLUDED IN THE MODEL	R	NP_VP	0.0	0.0	0.0	not	8.0	35.6	**7.6**	**<.01**	0.0	37.2	**24.4**	**<0.0001**	0.0	2.1	1.2	not
INCLUDED IN THIS STUDY	R	NP_VP	0.0	0.0	0.0	not	5.3	15.6	2.1	not	0.0	3.7	2.4	not	2.5	12.6	3.2	not
INCREASE IN THE AMOUNT OF	R	NP_VP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	13.2	1.9	**4.9**	**<.05**	2.5	2.1	0.0	not
INCREASE IN THE NUMBER OF	R	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	4.4	9.3	0.9	not	9.9	2.1	2.4	not
INCREASE THE LIKELIHOOD OF	R	NP_VP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	2.2	5.6	0.7	not	0.0	8.4	**4.9**	**<.05**
INDIVIDUALS ARE MORE LIKELY TO	P	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	0.0	0.0	not
INTERPRETATION OF THE RESULTS	T	NP	9.3	3.8	0.6	not	0.0	2.2	1.2	not	4.4	14.9	2.9	not	4.9	6.3	0.1	not
INVESTIGATE THE EFFECT OF	R	NP_VP	18.5	3.8	2.6	not	8.0	0.0	**4.7**	**<.05**	11.0	1.9	3.6	not	7.4	2.1	1.4	not
IS A FUNCTION OF	T	NP_VP	13.9	3.8	1.5	not	13.3	4.5	1.9	not	24.3	3.7	**8.5**	**<.01**	9.9	6.3	0.3	not
IS A FUNCTION OF THE	T	NP_VP	13.9	3.8	1.5	not	8.0	4.5	0.4	not	13.2	1.9	**4.9**	**<.05**	9.9	0.0	**6.2**	**<.05**
IS A MEASURE OF	R	NP_VP	0.0	3.8	1.2	not	18.6	15.6	0.1	not	4.4	1.9	0.5	not	0.0	0.0	0.0	not

Phrase																		
IS A MEASURE OF THE	R	NP_VP	0.0	0.0	0.0	not	10.6	4.5	1.1	not	2.2	1.9	0.0	not	0.0	0.0	0.0	not
IS ALSO CONSISTENT WITH	T	NP_VP	0.0	0.0	0.0	not	2.7	6.7	0.7	not	13.2	1.9	**4.9**	**<.05**	2.5	10.5	2.3	not
IS ALSO INTERESTING TO NOTE THAT	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	3.7	0.2	not	2.5	6.3	0.8	not
IS BASED ON A	T	VP	18.5	18.8	0.0	not	0.0	13.4	**7.3**	**<.01**	0.0	5.6	3.7	not	7.4	16.9	1.6	not
IS BASED ON THE	T	VP	23.1	22.6	0.0	not	18.6	62.4	**10.0**	**<.01**	4.4	9.3	0.9	not	17.3	14.8	0.1	not
IS BEYOND THE SCOPE	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	0.0	1.6	not	17.3	19.0	0.0	not
IS BEYOND THE SCOPE OF THIS	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	0.0	1.6	not	7.4	16.9	1.6	not
IS CHARACTERIZED BY A	R	VP	4.6	3.8	0.0	not	8.0	2.2	1.4	not	0.0	5.6	3.7	not	2.5	8.4	1.5	not
IS CONSISTENT WITH A	T	VP	4.6	11.3	0.7	not	2.7	4.5	0.2	not	11.0	1.9	3.6	not	14.8	4.2	2.8	not
IS CONSISTENT WITH PREVIOUS	T	VP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	11.0	1.9	3.6	not	19.8	10.5	1.3	not
IS CONSISTENT WITH PREVIOUS STUDIES	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	4.4	0.0	3.1	not	14.8	2.1	**4.8**	**<.05**
IS CONSISTENT WITH THE	T	VP	4.6	3.8	0.0	not	8.0	13.4	0.6	not	116.9	35.3	**23.1**	**<0.0001**	101.3	46.4	**9.2**	**<.01**
IS DEFINED AS THE	T	VP	13.9	30.1	1.4	not	21.3	20.0	0.0	not	26.5	0.0	**18.8**	**<0.0001**	7.4	0.0	**4.7**	**<.05**
IS DEPENDENT ON THE	T	VP	18.5	15.1	0.1	not	2.7	0.0	1.6	not	2.2	0.0	1.6	not	12.3	6.3	0.9	not
IS DETERMINED BY THE	R	VP	9.3	3.8	0.6	not	8.0	4.5	0.4	not	15.4	1.9	**6.2**	**<.05**	17.3	4.2	3.8	not
IS DUE TO THE	T	VP	23.1	3.8	3.8	not	8.0	4.5	0.4	not	52.9	11.1	**14.9**	**<.001**	49.4	14.8	**8.8**	**<.01**
IS EXPECTED TO BE	T	VP	23.1	0.0	**8.0**	**<.01**	0.0	0.0	0.0	not	11.0	7.4	0.3	not	24.7	0.0	**15.5**	**<0.0001**
IS FOUND TO BE	P	VP	4.6	3.8	0.0	not	2.7	0.0	1.6	not	13.2	14.9	0.0	not	14.8	6.3	1.6	not
IS IMPORTANT TO NOTE	P	VP	4.6	15.1	1.4	not	2.7	17.8	**5.0**	**<.05**	6.6	39.0	**12.3**	**<.001**	17.3	31.6	1.8	not
IS IMPORTANT TO NOTE THAT THE	P	VP	0.0	7.5	2.4	not	2.7	2.2	0.0	not	2.2	9.3	2.3	not	7.4	10.5	0.2	not
IS IN ACCORDANCE WITH THE	P	NP_VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	2.2	1.9	0.0	not	14.8	4.2	2.8	not
IS IN GOOD AGREEMENT WITH	T	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	13.2	0.0	**9.4**	**<.01**	19.8	0.0	**12.4**	**<.001**
IS IN GOOD AGREEMENT WITH THE	T	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	8.8	0.0	**6.3**	**<.05**	7.4	0.0	**4.7**	**<.05**
IS IN LINE WITH	T	NP_VP	9.3	3.8	0.6	not	2.7	11.1	2.2	not	2.2	44.6	**22.5**	**<0.0001**	17.3	42.2	**4.6**	**<.05**
IS IN LINE WITH THE	T	NP_VP	9.3	3.8	0.6	not	2.7	0.0	1.6	not	2.2	14.9	**5.1**	**<.05**	7.4	10.5	0.2	not
IS INTERESTING TO NOTE	P	VP	0.0	3.8	1.2	not	2.7	0.0	1.6	not	11.0	13.0	0.1	not	7.4	10.5	0.2	not
IS INTERESTING TO NOTE THAT THE	P	VP	0.0	3.8	1.2	not	2.7	0.0	1.6	not	4.4	0.0	3.1	not	4.9	4.2	0.0	not
IS LARGER THAN THAT OF THE	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	3.7	0.0	not	0.0	0.0	0.0	not
IS LESS LIKELY TO	P	VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	4.4	9.3	0.9	not	0.0	12.6	**7.4**	**<.01**
IS LIKELY TO BE	P	VP	23.1	11.3	1.0	not	2.7	4.5	0.2	not	13.2	18.6	0.4	not	32.1	52.7	2.2	not
IS LIKELY TO HAVE	P	VP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	9.9	4.2	1.0	not
IS MORE LIKELY TO	P	VP	0.0	30.1	**9.5**	**<.01**	0.0	13.4	**7.3**	**<.01**	2.2	24.2	**10.2**	**<.01**	9.9	27.4	3.7	not
IS MORE LIKELY TO BE	P	VP	0.0	11.3	3.6	not	0.0	2.2	1.2	not	2.2	11.1	3.2	not	2.5	8.4	1.5	not
IS MOST LIKELY TO	P	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	8.8	7.4	0.1	not	2.5	21.1	**7.2**	**<.01**
IS NEGATIVELY ASSOCIATED WITH	R	VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	2.2	24.2	**10.2**	**<.01**	2.5	8.4	1.5	not
IS NEGATIVELY RELATED TO	R	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	11.1	**7.3**	**<.01**	0.0	10.5	**6.2**	**<.05**
IS NOT THE CASE	T	NP_VP	4.6	7.5	0.2	not	2.7	2.2	0.0	not	8.8	20.4	2.3	not	14.8	12.6	0.1	not
IS NOT THE CASE FOR	T	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	6.6	7.4	0.0	not	7.4	2.1	1.4	not
IS ONE OF THE	P	NP_VP	148.0	41.4	**15.5**	**<0.0001**	8.0	26.7	**4.3**	**<.05**	2.2	22.3	**9.2**	**<.01**	29.6	23.2	0.3	not
IS ONE OF THE MOST	P	NP_VP	78.6	18.8	**9.6**	**<.01**	2.7	13.4	3.1	not	0.0	3.7	2.4	not	9.9	6.3	0.3	not
IS ONE OF THE MOST IMPORTANT	P	NP_VP	23.1	3.8	3.8	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	7.4	0.0	**4.7**	**<.05**
IS POSITIVE AND SIGNIFICANT	R	VP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	0.0	44.6	**29.3**	**<0.0001**	0.0	4.2	2.5	not
IS POSITIVELY ASSOCIATED WITH	R	VP	0.0	7.5	2.4	not	0.0	8.9	**4.9**	**<.05**	0.0	29.7	**19.6**	**<0.0001**	0.0	10.5	**6.2**	**<.05**
IS POSITIVELY RELATED TO	R	VP	0.0	30.1	**9.5**	**<.01**	2.7	2.2	0.0	not	0.0	40.9	**26.9**	**<0.0001**	0.0	14.8	**8.6**	**<.01**
IS POSSIBLE THAT THE	P	VP	4.6	7.5	0.2	not	2.7	2.2	0.0	not	6.6	11.1	0.6	not	46.9	35.8	0.7	not
IS PROPORTIONAL TO THE	R	VP	13.9	0.0	**4.8**	**<.05**	10.6	2.2	2.5	not	22.1	0.0	**15.7**	**<0.0001**	9.9	0.0	**6.2**	**<.05**
IS SHOWN IN FIG	T	NP_VP	9.3	3.8	0.6	not	111.8	11.1	**40.3**	**<0.0001**	185.3	1.9	**121.8**	**<0.0001**	42.0	0.0	**26.4**	**<0.0001**
IS SHOWN IN FIGURE	T	NP_VP	0.0	0.0	0.0	not	13.3	6.7	0.9	not	44.1	1.9	**24.5**	**<0.0001**	4.9	0.0	3.1	not
IS SIGNIFICANT AT THE	R	VP	0.0	0.0	0.0	not	2.7	11.1	2.2	not	2.2	24.2	**10.2**	**<.01**	0.0	0.0	0.0	not

IS SIMILAR TO THAT	T	VP	23.1	0.0	**8.0**	**<.01**	16.0	2.2	**4.9**	**<.05**	19.9	3.7	**6.1**	**<.05**	27.2	4.2	**8.4**	**<.01**
IS SIMILAR TO THAT OF	T	NP_VP	23.1	0.0	**8.0**	**<.01**	5.3	0.0	3.1	not	6.6	0.0	**4.7**	**<.05**	12.3	2.1	3.6	not
IS SIMILAR TO THE	T	VP	4.6	7.5	0.2	not	10.6	20.0	1.2	not	17.6	1.9	**7.5**	**<.01**	17.3	10.5	0.7	not
IS SUPPORTED BY THE	T	VP	4.6	11.3	0.7	not	0.0	2.2	1.2	not	8.8	3.7	1.1	not	17.3	6.3	2.3	not
IS THE EXTENT TO WHICH	R	NP_VP	0.0	3.8	1.2	not	0.0	8.9	**4.9**	**<.05**	0.0	9.3	**6.1**	**<.05**	7.4	0.0	**4.7**	**<.05**
IS THE NUMBER OF	R	NP_VP	4.6	3.8	0.0	not	42.6	13.4	**6.7**	**<.01**	4.4	3.7	0.0	not	7.4	4.2	0.4	not
IS THE RESULT OF	T	NP_VP	4.6	7.5	0.2	not	0.0	4.5	2.4	not	13.2	11.1	0.1	not	12.3	19.0	0.6	not
IS THE RESULT OF A	T	NP_VP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	8.8	3.7	1.1	not	0.0	0.0	0.0	not
IS THE SAME AS	T	NP_VP	4.6	3.8	0.0	not	10.6	0.0	**6.3**	**<.05**	17.6	5.6	3.3	not	2.5	2.1	0.0	not
IS THOUGHT TO BE	P	VP	9.3	7.5	0.0	not	5.3	0.0	3.1	not	11.0	0.0	**7.8**	**<.01**	12.3	0.0	**7.8**	**<.01**
IS UNLIKELY TO BE	P	VP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	0.0	3.7	2.4	not	12.3	6.3	0.9	not
IS WORTH NOTING THAT	P	VP	18.5	0.0	**6.4**	**<.05**	24.0	8.9	3.0	not	15.4	16.7	0.0	not	9.9	21.1	1.8	not
IT APPEARS THAT THE	P	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	11.0	20.4	1.4	not	24.7	27.4	0.1	not
IT CAN BE ARGUED	P	VP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	21.1	**12.3**	**<.001**
IT CAN BE ARGUED THAT	P	VP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	0.0	21.1	**12.3**	**<.001**
IT CAN BE ARGUED THAT THE	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	12.6	**7.4**	**<.01**
IT CAN BE CONCLUDED THAT	P	VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	26.5	5.6	**7.4**	**<.01**	37.0	6.3	**10.7**	**<.01**
IT CAN BE CONCLUDED THAT THE	P	VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	15.4	1.9	**6.2**	**<.05**	22.2	4.2	**6.0**	**<.05**
IT CAN BE OBSERVED	P	VP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	35.3	1.9	**18.7**	**<0.0001**	2.5	0.0	1.6	not
IT CAN BE OBSERVED THAT	T	VP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	30.9	1.9	**15.8**	**<0.0001**	2.5	0.0	1.6	not
IT CAN BE OBSERVED THAT THE	P	VP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	19.9	0.0	**14.1**	**<.001**	0.0	0.0	0.0	not
IT CAN BE SEEN	T	VP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	70.6	18.6	**16.2**	**<0.0001**	22.2	2.1	**8.7**	**<.01**
IT CAN BE SEEN FROM FIG	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	17.6	0.0	**12.5**	**<.001**	0.0	0.0	0.0	not
IT CAN BE SEEN THAT	T	VP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	41.9	13.0	**8.0**	**<.01**	22.2	2.1	**8.7**	**<.01**
IT CAN BE SEEN THAT THE	T	VP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	26.5	5.6	**7.4**	**<.01**	14.8	0.0	**9.3**	**<.01**
IT COULD BE ARGUED	P	VP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	2.2	11.1	3.2	not	4.9	14.8	2.2	not
IT COULD BE ARGUED THAT	P	VP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	2.2	11.1	3.2	not	4.9	14.8	2.2	not
IT HAS ALSO BEEN	T	VP	37.0	3.8	**7.7**	**<.01**	0.0	6.7	3.6	not	0.0	1.9	1.2	not	9.9	6.3	0.3	not
IT HAS BEEN ARGUED THAT	T	VP	0.0	26.4	**8.3**	**<.01**	0.0	2.2	1.2	not	0.0	0.0	0.0	not	2.5	8.4	1.5	not
IT HAS BEEN DEMONSTRATED THAT	P	VP	23.1	0.0	**8.0**	**<.01**	0.0	0.0	0.0	not	4.4	0.0	3.1	not	4.9	2.1	0.5	not
IT HAS BEEN REPORTED THAT	T	VP	32.4	0.0	**11.2**	**<.001**	0.0	0.0	0.0	not	11.0	0.0	**7.8**	**<.01**	29.6	0.0	**18.6**	**<0.0001**
IT HAS BEEN SHOWN	T	VP	78.6	3.8	**20.7**	**<0.0001**	5.3	2.2	0.5	not	24.3	3.7	**8.5**	**<.01**	61.7	6.3	**23.4**	**<0.0001**
IT HAS BEEN SHOWN THAT	T	VP	60.1	3.8	**14.8**	**<.001**	5.3	2.2	0.5	not	24.3	3.7	**8.5**	**<.01**	46.9	6.3	**15.6**	**<0.0001**
IT HAS BEEN SHOWN THAT THE	T	VP	9.3	0.0	3.2	not	5.3	2.2	0.5	not	2.2	0.0	1.6	not	9.9	0.0	**6.2**	**<.05**
IT HAS BEEN SUGGESTED THAT	T	VP	37.0	11.3	3.5	not	0.0	2.2	1.2	not	2.2	7.4	1.4	not	24.7	0.0	**15.5**	**<0.0001**
IT HAS BEEN SUGGESTED THAT THE	T	VP	9.3	0.0	3.2	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	7.4	0.0	**4.7**	**<.05**
IT IS ALSO IMPORTANT	P	VP	4.6	3.8	0.0	not	0.0	2.2	1.2	not	4.4	9.3	0.9	not	9.9	12.6	0.1	not
IT IS ALSO IMPORTANT TO	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	2.2	9.3	2.3	not	7.4	10.5	0.2	not
IT IS ALSO INTERESTING TO NOTE THAT	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	3.7	0.2	not	2.5	6.3	0.8	not
IT IS ALSO POSSIBLE THAT	P	VP	0.0	11.3	3.6	not	0.0	8.9	**4.9**	**<.05**	2.2	7.4	1.4	not	19.8	25.3	0.3	not
IT IS ALSO WORTH NOTING THAT	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	3.7	0.0	not	2.5	0.0	1.6	not
IT IS APPARENT THAT	P	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	24.3	3.7	**8.5**	**<.01**	7.4	4.2	0.4	not
IT IS ARGUED THAT	P	VP	4.6	11.3	0.7	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	8.4	**4.9**	**<.05**
IT IS ASSUMED THAT	P	VP	0.0	11.3	3.6	not	10.6	11.1	0.0	not	11.0	3.7	1.9	not	9.9	2.1	2.4	not
IT IS ASSUMED THAT THE	P	VP	0.0	0.0	0.0	not	8.0	4.5	0.4	not	6.6	1.9	1.4	not	7.4	2.1	1.4	not
IT IS CLEAR THAT	P	VP	4.6	3.8	0.0	not	0.0	2.2	1.2	not	46.3	11.1	**11.6**	**<.001**	17.3	23.2	0.4	not
IT IS CLEAR THAT THE	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	26.5	5.6	**7.4**	**<.01**	12.3	6.3	0.9	not
IT IS DIFFICULT TO	P	VP	27.8	30.1	0.0	not	8.0	20.0	2.2	not	13.2	9.3	0.3	not	42.0	25.3	1.8	not
IT IS ESSENTIAL TO	P	VP	18.5	11.3	0.4	not	0.0	2.2	1.2	not	2.2	1.9	0.0	not	7.4	8.4	0.0	not

IT IS EVIDENT THAT	P	VP	0.0	3.8	1.2	not	2.7	2.2	0.0	not	22.1	3.7	**7.3**	**<.01**	17.3	6.3	2.3	not
IT IS EVIDENT THAT THE	P	VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	8.8	3.7	1.1	not	12.3	0.0	**7.8**	**<.01**
IT IS EXPECTED THAT	T	VP	9.3	7.5	0.0	not	2.7	2.2	0.0	not	4.4	1.9	0.5	not	19.8	6.3	3.2	not
IT IS EXPECTED THAT THE	P	VP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	4.4	0.0	3.1	not	4.9	4.2	0.0	not
IT IS FOUND THAT	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	5.6	1.6	not	17.3	0.0	**10.9**	**<.001**
IT IS FOUND THAT THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	1.9	0.5	not	9.9	0.0	**6.2**	**<.05**
IT IS IMPORTANT TO	P	VP	13.9	60.2	**7.3**	**<.01**	13.3	49.0	**8.7**	**<.01**	19.9	55.7	**8.6**	**<.01**	51.9	86.4	3.8	not
IT IS IMPORTANT TO CONSIDER	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	5.6	3.7	not	9.9	2.1	2.4	not
IT IS IMPORTANT TO NOTE	P	VP	4.6	15.1	1.4	not	2.7	17.8	**5.0**	**<.05**	6.6	35.3	**10.4**	**<.01**	17.3	29.5	1.4	not
IT IS IMPORTANT TO NOTE THAT	P	VP	4.6	11.3	0.7	not	2.7	15.6	**4.1**	**<.05**	6.6	22.3	**4.3**	**<.05**	14.8	25.3	1.2	not
IT IS IMPORTANT TO NOTE THAT THE	P	VP	0.0	7.5	2.4	not	2.7	2.2	0.0	not	2.2	9.3	2.3	not	7.4	10.5	0.2	not
IT IS IMPORTANT TO UNDERSTAND	P	VP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	10.5	2.3	not
IT IS INTERESTING TO NOTE THAT	P	VP	0.0	3.8	1.2	not	2.7	0.0	1.6	not	11.0	7.4	0.3	not	4.9	6.3	0.1	not
IT IS INTERESTING TO NOTE THAT THE	P	VP	0.0	3.8	1.2	not	2.7	0.0	1.6	not	4.4	0.0	3.1	not	4.9	4.2	0.0	not
IT IS KNOWN THAT	P	VP	9.3	7.5	0.0	not	0.0	0.0	0.0	not	17.6	0.0	**12.5**	**<.001**	27.2	0.0	**17.1**	**<0.0001**
IT IS LIKELY THAT	P	VP	13.9	0.0	**4.8**	**<.05**	2.7	13.4	3.1	not	15.4	3.7	**3.9**	**<.05**	44.5	56.9	0.7	not
IT IS LIKELY THAT THE	P	VP	4.6	0.0	1.6	not	0.0	4.5	2.4	not	4.4	1.9	0.5	not	24.7	25.3	0.0	not
IT IS MORE LIKELY THAT	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	9.9	4.2	1.0	not
IT IS NECESSARY TO	P	VP	55.5	15.1	**6.0**	**<.05**	2.7	17.8	**5.0**	**<.05**	4.4	20.4	**5.4**	**<.05**	22.2	12.6	1.2	not
IT IS NOT CLEAR	P	VP	13.9	22.6	0.5	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	14.8	8.4	0.8	not
IT IS NOT POSSIBLE TO	P	VP	9.3	0.0	3.2	not	0.0	6.7	3.6	not	11.0	11.1	0.0	not	7.4	10.5	0.2	not
IT IS NOT SURPRISING	P	VP	9.3	7.5	0.0	not	0.0	0.0	0.0	not	4.4	11.1	1.5	not	2.5	10.5	2.3	not
IT IS NOT SURPRISING THAT	P	VP	9.3	7.5	0.0	not	0.0	0.0	0.0	not	2.2	9.3	2.3	not	0.0	10.5	**6.2**	**<.05**
IT IS NOTED THAT	P	VP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	19.9	1.9	**8.8**	**<.01**	24.7	0.0	**15.5**	**<0.0001**
IT IS POSSIBLE THAT	P	VP	18.5	33.9	1.1	not	2.7	13.4	3.1	not	26.5	29.7	0.1	not	123.5	122.2	0.0	not
IT IS POSSIBLE THAT SOME	P	VP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	4.9	6.3	0.1	not
IT IS POSSIBLE THAT THE	P	VP	4.6	7.5	0.2	not	2.7	2.2	0.0	not	6.6	11.1	0.6	not	46.9	35.8	0.7	not
IT IS POSSIBLE TO	P	VP	37.0	22.6	0.8	not	10.6	6.7	0.4	not	13.2	13.0	0.0	not	19.8	14.8	0.3	not
IT IS REASONABLE TO	P	VP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	15.4	1.9	**6.2**	**<.05**	22.2	6.3	**4.2**	**<.05**
IT IS REASONABLE TO ASSUME THAT	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	2.2	0.0	1.6	not	2.5	2.1	0.0	not
IT IS UNLIKELY THAT	P	VP	0.0	0.0	0.0	not	0.0	8.9	**4.9**	**<.05**	8.8	3.7	1.1	not	22.2	14.8	0.7	not
IT IS WELL KNOWN THAT	P	VP	13.9	0.0	**4.8**	**<.05**	2.7	2.2	0.0	not	4.4	0.0	3.1	not	14.8	2.1	**4.8**	**<.05**
IT IS WORTH NOTING	P	VP	18.5	0.0	**6.4**	**<.05**	24.0	8.9	3.0	not	15.4	16.7	0.0	not	12.3	21.1	1.0	not
IT IS WORTH NOTING THAT	P	VP	18.5	0.0	**6.4**	**<.05**	24.0	8.9	3.0	not	15.4	16.7	0.0	not	9.9	21.1	1.8	not
IT IS WORTH NOTING THAT THE	P	VP	0.0	0.0	0.0	not	5.3	4.5	0.0	not	8.8	1.9	2.5	not	2.5	10.5	2.3	not
IT MAY BE THAT	P	VP	0.0	3.8	1.2	not	0.0	8.9	**4.9**	**<.05**	4.4	1.9	0.5	not	12.3	59.0	**14.2**	**<.001**
IT SHOULD ALSO BE	P	VP	0.0	0.0	0.0	not	5.3	4.5	0.0	not	17.6	3.7	**5.0**	**<.05**	4.9	6.3	0.1	not
IT SHOULD ALSO BE NOTED THAT	P	VP	0.0	0.0	0.0	not	2.7	4.5	0.2	not	15.4	3.7	**3.9**	**<.05**	4.9	2.1	0.5	not
IT SHOULD BE NOTED	P	VP	23.1	7.5	2.0	not	45.2	26.7	2.0	not	81.6	35.3	**9.4**	**<.01**	49.4	27.4	2.8	not
IT SHOULD BE NOTED HOWEVER THAT	P	VP	4.6	3.8	0.0	not	2.7	2.2	0.0	not	0.0	1.9	1.2	not	2.5	2.1	0.0	not
IT SHOULD BE NOTED THAT	P	VP	18.5	3.8	2.6	not	42.6	22.3	2.7	not	81.6	29.7	**12.5**	**<.001**	44.5	23.2	3.0	not
IT SHOULD BE NOTED THAT THE	P	VP	9.3	0.0	3.2	not	21.3	6.7	3.3	not	33.1	14.9	3.5	not	27.2	8.4	**4.6**	**<.05**
IT SHOULD BE NOTED THAT THIS	P	VP	4.6	0.0	1.6	not	8.0	0.0	**4.7**	**<.05**	2.2	0.0	1.6	not	0.0	2.1	1.2	not
IT WAS ALSO FOUND THAT THE	P	VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	4.4	0.0	3.1	not	2.5	4.2	0.2	not
IT WAS FOUND THAT	P	VP	46.3	0.0	**16.0**	**<0.0001**	13.3	2.2	3.7	not	33.1	9.3	**7.1**	**<.01**	29.6	10.5	**4.2**	**<.05**
IT WAS FOUND THAT THE	T	VP	9.3	0.0	3.2	not	8.0	2.2	1.4	not	15.4	3.7	**3.9**	**<.05**	7.4	4.2	0.4	not
IT WAS NOT POSSIBLE TO	P	VP	0.0	0.0	0.0	not	10.6	8.9	0.1	not	6.6	5.6	0.0	not	9.9	0.0	**6.2**	**<.05**
IT WOULD BE INTERESTING	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	12.3	29.5	3.1	not
IT WOULD BE INTERESTING TO	P	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	12.3	25.3	2.0	not

LARGER THAN THAT OF	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	11.1	0.1	not	9.9	2.1	2.4	not
LEAD TO THE FORMATION OF	R	NP_VP	13.9	0.0	**4.8**	**<.05**	0.0	2.2	1.2	not	2.2	0.0	1.6	not	9.9	0.0	**6.2**	**<.05**
LESS ATTENTION HAS BEEN PAID TO	T	VP	0.0	22.6	**7.1**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
LESS LIKELY TO BE	P	VP	4.6	0.0	1.6	not	0.0	8.9	**4.9**	**<.05**	6.6	9.3	0.2	not	2.5	25.3	**9.3**	**<.01**
LESS LIKELY TO ENGAGE IN	P	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	6.3	3.7	not
LIES IN THE FACT THAT	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	7.4	4.2	0.4	not
LIKELY THAN MEN TO	P	PP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	11.1	**7.3**	**<.01**	0.0	6.3	3.7	not
LIKELY TO BE DUE TO	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	15.4	1.9	**6.2**	**<.05**	4.9	0.0	3.1	not
LIKELY TO HAVE A	P	VP	0.0	3.8	1.2	not	0.0	8.9	**4.9**	**<.05**	2.2	9.3	2.3	not	12.3	10.5	0.1	not
LIMITATION OF THIS STUDY IS	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	7.4	16.9	1.6	not
LITTLE IS KNOWN ABOUT	T	VP	64.8	48.9	0.5	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	9.9	10.5	0.0	not
LITTLE IS KNOWN ABOUT HOW	T	VP	9.3	11.3	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	6.3	0.8	not
LITTLE IS KNOWN ABOUT THE	T	VP	46.3	18.8	2.9	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	2.5	4.2	0.2	not
LOWER THAN THAT OF THE	R	PP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	17.6	3.7	**5.0**	**<.05**	22.2	4.2	**6.0**	**<.05**
MAKES IT POSSIBLE TO	P	VP	4.6	11.3	0.7	not	0.0	17.8	**9.7**	**<.01**	0.0	3.7	2.4	not	7.4	2.1	1.4	not
MAY BE ABLE TO	P	VP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	2.2	1.9	0.0	not	7.4	16.9	1.6	not
MAY BE ATTRIBUTED TO	P	VP	4.6	3.8	0.0	not	0.0	2.2	1.2	not	4.4	0.0	3.1	not	24.7	8.4	3.7	not
MAY BE DUE TO	T	VP	18.5	3.8	2.6	not	0.0	4.5	2.4	not	30.9	18.6	1.5	not	66.7	21.1	**11.0**	**<.001**
MAY BE DUE TO THE	T	VP	9.3	0.0	3.2	not	0.0	2.2	1.2	not	8.8	11.1	0.1	not	37.0	16.9	3.4	not
MAY BE EXPLAINED BY	P	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	4.4	9.3	0.9	not	17.3	14.8	0.1	not
MAY BE EXPLAINED BY THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	5.6	0.1	not	2.5	12.6	3.2	not
MAY BE INFLUENCED BY	P	VP	4.6	18.8	2.2	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	2.5	8.4	1.5	not
MAY BE LESS LIKELY TO	P	VP	4.6	3.8	0.0	not	0.0	2.2	1.2	not	0.0	1.9	1.2	not	0.0	10.5	**6.2**	**<.05**
MAY BE MORE LIKELY	P	VP	0.0	18.8	**6.0**	**<.05**	0.0	6.7	3.6	not	0.0	11.1	**7.3**	**<.01**	4.9	14.8	2.2	not
MAY BE MORE LIKELY TO	P	VP	0.0	15.1	**4.8**	**<.05**	0.0	6.7	3.6	not	0.0	9.3	**6.1**	**<.05**	4.9	12.6	1.5	not
MAY BE RELATED TO THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	11.0	1.9	3.6	not	12.3	8.4	0.3	not
MAY NOT BE ABLE TO	P	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.9	6.3	0.1	not
MAY OR MAY NOT	P	VP	0.0	3.8	1.2	not	2.7	4.5	0.2	not	0.0	0.0	0.0	not	0.0	16.9	**9.9**	**<.01**
MAY PLAY A ROLE IN	P	VP	13.9	0.0	**4.8**	**<.05**	0.0	0.0	0.0	not	2.2	0.0	1.6	not	9.9	4.2	1.0	not
MEAN SQUARE ERROR OF	R	NP	0.0	0.0	0.0	not	0.0	17.8	**9.7**	**<.01**	2.2	18.6	**7.1**	**<.01**	0.0	0.0	0.0	not
MEAN SQUARE ERROR OF APPROXIMATION	R	NP	0.0	0.0	0.0	not	0.0	17.8	**9.7**	**<.01**	0.0	18.6	**12.2**	**<.001**	0.0	0.0	0.0	not
MEASURED AS THE NUMBER OF	R	NP_VP	0.0	0.0	0.0	not	2.7	20.0	**6.0**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
MEASURED BY THE NUMBER OF	R	NP_VP	0.0	3.8	1.2	not	0.0	22.3	**12.2**	**<.001**	0.0	0.0	0.0	not	0.0	2.1	1.2	not
MECHANICAL BEHAVIOR OF THE	R	NP	9.3	0.0	3.2	not	10.6	0.0	**6.3**	**<.05**	17.6	0.0	**12.5**	**<.001**	12.3	0.0	**7.8**	**<.01**
MECHANICAL PROPERTIES OF THE	R	NP	27.8	0.0	**9.6**	**<.01**	24.0	0.0	**14.1**	**<.001**	17.6	0.0	**12.5**	**<.001**	32.1	0.0	**20.2**	**<0.0001**
MEDIATE THE RELATIONSHIP BETWEEN	T	NP_VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	7.4	**4.9**	**<.05**	0.0	4.2	2.5	not
MEN ARE MORE LIKELY TO	P	NP_VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	26.0	**17.1**	**<0.0001**	0.0	8.4	**4.9**	**<.05**
MIGHT BE DUE TO	T	VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	11.0	9.3	0.1	not	14.8	8.4	0.8	not
MIGHT BE DUE TO THE	T	VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	11.0	5.6	0.9	not	7.4	4.2	0.4	not
MIGHT BE MORE LIKELY TO	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	7.4	**4.9**	**<.05**	0.0	8.4	**4.9**	**<.05**
MIGHT BE RELATED TO THE	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	6.6	5.6	0.0	not	2.5	6.3	0.8	not
MIN AT ROOM TEMPERATURE	R	PP	0.0	0.0	0.0	not	53.2	0.0	**31.4**	**<0.0001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
MORE LIKELY THAN MEN	P	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	11.1	**7.3**	**<.01**	0.0	2.1	1.2	not
MORE LIKELY THAN MEN TO	P	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	9.3	**6.1**	**<.05**	0.0	2.1	1.2	not
MORE LIKELY THAN WOMEN TO	P	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	5.6	3.7	not	0.0	6.3	3.7	not
MORE LIKELY TO BE	P	PP	0.0	48.9	**15.5**	**<0.0001**	2.7	26.7	**9.1**	**<.01**	6.6	96.6	**45.0**	**<0.0001**	9.9	71.7	**22.6**	**<0.0001**
MORE LIKELY TO ENGAGE	P	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	16.7	**11.0**	**<.001**	0.0	6.3	3.7	not
MORE LIKELY TO ENGAGE IN	P	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	14.9	**9.8**	**<.01**	0.0	6.3	3.7	not
MORE LIKELY TO HAVE	P	PP	4.6	11.3	0.7	not	0.0	13.4	**7.3**	**<.01**	6.6	31.6	**8.6**	**<.01**	0.0	21.1	**12.3**	**<.001**

Phrase																		
MORE OR LESS LIKELY TO	P	PP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
MORE THAN HALF OF	R	PP	4.6	0.0	1.6	not	5.3	6.7	0.1	not	8.8	22.3	2.9	not	4.9	12.6	1.5	not
MORE THAN HALF OF THE	R	PP	4.6	0.0	1.6	not	2.7	4.5	0.2	not	6.6	16.7	2.2	not	2.5	6.3	0.8	not
MOST LIKELY TO BE	P	PP	0.0	3.8	1.2	not	2.7	2.2	0.0	not	17.6	13.0	0.3	not	2.5	12.6	3.2	not
MUCH MORE LIKELY TO	P	PP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	2.2	18.6	**7.1**	**<.01**	0.0	8.4	**4.9**	**<.05**
N IS THE NUMBER OF	R	NP_VP	4.6	0.0	1.6	not	18.6	8.9	1.4	not	0.0	0.0	0.0	not	2.5	0.0	1.6	not
NO SIGNIFICANT DIFFERENCE BETWEEN	R	NP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	19.9	16.7	0.1	not	2.5	0.0	1.6	not
NO SIGNIFICANT DIFFERENCE IN	R	NP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	26.5	7.4	**5.7**	**<.05**	9.9	2.1	2.4	not
NO SIGNIFICANT DIFFERENCES BETWEEN THE	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	13.0	0.0	not	2.5	0.0	1.6	not
NO SIGNIFICANT DIFFERENCES WERE FOUND	R	NP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	8.8	9.3	0.0	not	0.0	0.0	0.0	not
NO STATISTICALLY SIGNIFICANT DIFFERENCES	R	NP	0.0	0.0	0.0	not	0.0	8.9	**4.9**	**<.05**	24.3	9.3	3.5	not	9.9	0.0	**6.2**	**<.05**
NOT APPEAR TO BE	P	VP	4.6	18.8	2.2	not	2.7	8.9	1.4	not	8.8	7.4	0.1	not	17.3	0.0	**10.9**	**<.001**
NOT BE ABLE TO	P	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	7.4	10.5	0.2	not
NOT HAVE A SIGNIFICANT EFFECT ON	R	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	4.4	7.4	0.4	not	0.0	0.0	0.0	not
NOT SIGNIFICANTLY DIFFERENT FROM	R	OTHER	0.0	0.0	0.0	not	0.0	8.9	**4.9**	**<.05**	15.4	20.4	0.3	not	2.5	2.1	0.0	not
OF A LARGE NUMBER OF	R	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	9.9	4.2	1.0	not
OF A WIDE RANGE OF	R	PP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	2.2	0.0	1.6	not	4.9	4.2	0.0	not
OF DIFFERENT TYPES OF	T	PP	4.6	11.3	0.7	not	2.7	4.5	0.2	not	0.0	11.1	**7.3**	**<.01**	4.9	10.5	0.9	not
OF EACH OF THE	R	PP	9.3	0.0	3.2	not	24.0	20.0	0.1	not	8.8	14.9	0.8	not	9.9	4.2	1.0	not
OF THE CURRENT STUDY	T	PP	13.9	26.4	0.9	not	2.7	2.2	0.0	not	2.2	1.9	0.0	not	14.8	44.3	**6.6**	**<.05**
OF THE CURRENT STUDY IS TO	T	NP_VP	0.0	15.1	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
OF THE DEPENDENT VARIABLE	R	PP	0.0	3.8	1.2	not	0.0	31.2	**17.0**	**<0.0001**	0.0	35.3	**23.2**	**<0.0001**	0.0	10.5	**6.2**	**<.05**
OF THE EFFECT OF	R	PP	32.4	7.5	**4.1**	**<.05**	8.0	20.0	2.2	not	2.2	22.3	**9.2**	**<.01**	9.9	4.2	1.0	not
OF THE EFFECTS OF	R	PP	9.3	33.9	3.5	not	2.7	6.7	0.7	not	6.6	18.6	2.9	not	9.9	31.6	**5.2**	**<.05**
OF THE EXTENT TO WHICH	R	PP	4.6	7.5	0.2	not	0.0	8.9	**4.9**	**<.05**	0.0	0.0	0.0	not	0.0	6.3	3.7	not
OF THE IMPACT OF	R	PP	23.1	18.8	0.1	not	5.3	4.5	0.0	not	2.2	7.4	1.4	not	4.9	16.9	3.0	not
OF THE IMPORTANCE OF	P	PP	4.6	7.5	0.2	not	0.0	4.5	2.4	not	2.2	9.3	2.3	not	7.4	12.6	0.6	not
OF THE INDEPENDENT VARIABLES	R	PP	0.0	3.8	1.2	not	2.7	33.4	**12.3**	**<.001**	0.0	20.4	**13.4**	**<.001**	0.0	0.0	0.0	not
OF THE INFLUENCE OF	R	PP	9.3	11.3	0.0	not	8.0	8.9	0.0	not	0.0	1.9	1.2	not	7.4	10.5	0.2	not
OF THE INTERACTION BETWEEN	T	PP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	2.2	20.4	**8.1**	**<.01**	12.3	4.2	1.8	not
OF THE MOST IMPORTANT	P	PP	37.0	26.4	0.4	not	0.0	6.7	3.6	not	4.4	3.7	0.0	not	9.9	10.5	0.0	not
OF THE NUMBER OF	R	PP	0.0	3.8	1.2	not	21.3	37.9	1.9	not	19.9	24.2	0.2	not	12.3	4.2	1.8	not
OF THE PRESENT STUDY IS TO	T	NP_VP	18.5	3.8	2.6	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
OF THE QUALITY OF	R	PP	0.0	22.6	**7.1**	**<.01**	0.0	6.7	3.6	not	2.2	3.7	0.2	not	4.9	4.2	0.0	not
OF THE RELATIONSHIP BETWEEN	T	PP	9.3	41.4	**5.1**	**<.05**	2.7	2.2	0.0	not	8.8	13.0	0.4	not	12.3	67.4	**17.9**	**<0.0001**
OF THE ROLE OF	R	PP	13.9	22.6	0.5	not	0.0	4.5	2.4	not	2.2	3.7	0.2	not	4.9	12.6	1.5	not
OF THE STUDY WAS	T	NP_VP	4.6	3.8	0.0	not	2.7	4.5	0.2	not	4.4	7.4	0.4	not	12.3	6.3	0.9	not
OF THE SUPPORTING INFORMATION	T	PP	4.6	0.0	1.6	not	24.0	0.0	**14.1**	**<.001**	24.3	0.0	**17.2**	**<0.0001**	32.1	0.0	**20.2**	**<0.0001**
OF THE TOTAL NUMBER OF	R	PP	0.0	0.0	0.0	not	8.0	17.8	1.6	not	0.0	7.4	**4.9**	**<.05**	0.0	2.1	1.2	not
OF THE VARIANCE IN	R	PP	0.0	3.8	1.2	not	5.3	6.7	0.1	not	4.4	44.6	**18.4**	**<0.0001**	4.9	35.8	**11.3**	**<.001**
OF THE VARIANCE IN THE	R	PP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	0.0	20.4	**13.4**	**<.001**	0.0	8.4	**4.9**	**<.05**
OF THE VARIATION IN	R	PP	9.3	3.8	0.6	not	2.7	11.1	2.2	not	24.3	11.1	2.5	not	9.9	4.2	1.0	not
OF THIS ANALYSIS ARE PRESENTED IN	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	7.4	0.4	not	0.0	0.0	0.0	not
OF THIS ARTICLE IS	T	NP_VP	0.0	41.4	**13.1**	**<.001**	0.0	4.5	2.4	not	0.0	1.9	1.2	not	0.0	8.4	**4.9**	**<.05**
OF THIS STUDY ARE	T	NP_VP	9.3	15.1	0.3	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	7.4	14.8	1.1	not
OF THIS STUDY IS	T	NP_VP	50.9	60.2	0.2	not	5.3	15.6	2.1	not	2.2	0.0	1.6	not	19.8	61.1	**9.6**	**<.01**
OF THIS STUDY IS THAT IT	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.9	6.3	0.1	not
OF THIS STUDY WAS	T	NP_VP	97.1	7.5	**22.5**	**<0.0001**	0.0	2.2	1.2	not	2.2	3.7	0.2	not	9.9	29.5	**4.4**	**<.05**
OF THIS STUDY WAS TO INVESTIGATE	T	NP_VP	13.9	0.0	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	2.1	0.0	not

ON A SAMPLE OF	R	PP	0.0	18.8	**6.0**	**<.05**	0.0	8.9	**4.9**	**<.05**	2.2	1.9	0.0	not	0.0	12.6	**7.4**	**<.01**
ON BOTH SIDES OF THE	R	PP	0.0	0.0	0.0	not	5.3	0.0	3.1	not	8.8	0.0	**6.3**	**<.05**	4.9	6.3	0.1	not
ON THE BASIS OF	T	PP	64.8	41.4	1.2	not	47.9	98.0	**7.1**	**<.01**	35.3	53.9	1.9	not	64.2	56.9	0.2	not
ON THE BASIS OF A	T	PP	9.3	3.8	0.6	not	0.0	11.1	**6.1**	**<.05**	4.4	7.4	0.4	not	7.4	4.2	0.4	not
ON THE BASIS OF THE	T	PP	41.6	15.1	3.1	not	31.9	26.7	0.2	not	22.1	29.7	0.6	not	22.2	27.4	0.2	not
ON THE BASIS OF THEIR	T	PP	9.3	0.0	3.2	not	2.7	8.9	1.4	not	8.8	3.7	1.1	not	4.9	8.4	0.4	not
ON THE BASIS OF THESE	T	PP	4.6	7.5	0.2	not	2.7	0.0	1.6	not	2.2	7.4	1.4	not	4.9	2.1	0.5	not
ON THE EFFECTS OF	R	PP	18.5	41.4	2.1	not	10.6	8.9	0.1	not	0.0	0.0	0.0	not	4.9	21.1	**4.6**	**<.05**
ON THE EXTENT TO WHICH	R	PP	0.0	22.6	**7.1**	**<.01**	0.0	8.9	**4.9**	**<.05**	0.0	7.4	**4.9**	**<.05**	2.5	0.0	1.6	not
ON THE IMPACT OF	R	PP	9.3	37.6	**4.3**	**<.05**	0.0	6.7	3.6	not	2.2	5.6	0.7	not	2.5	14.8	**4.2**	**<.05**
ON THE LEVEL OF	T	PP	0.0	11.3	3.6	not	0.0	11.1	**6.1**	**<.05**	2.2	5.6	0.7	not	4.9	14.8	2.2	not
ON THE LIKELIHOOD OF	R	PP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	0.0	27.9	**18.3**	**<0.0001**	0.0	10.5	**6.2**	**<.05**
ON THE NATURE OF	R	PP	9.3	15.1	0.3	not	0.0	4.5	2.4	not	4.4	3.7	0.0	not	4.9	10.5	0.9	not
ON THE NATURE OF THE	R	PP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	4.4	1.9	0.5	not	2.5	4.2	0.2	not
ON THE NUMBER OF	R	PP	0.0	11.3	3.6	not	10.6	35.6	**5.7**	**<.05**	4.4	24.2	**7.2**	**<.01**	12.3	8.4	0.3	not
ON THE ONE HAND	T	PP	9.3	60.2	**9.7**	**<.01**	0.0	31.2	**17.0**	**<0.0001**	6.6	35.3	**10.4**	**<.01**	7.4	46.4	**13.5**	**<.001**
ON THE ONE HAND AND	T	PP	0.0	30.1	**9.5**	**<.01**	0.0	17.8	**9.7**	**<.01**	0.0	9.3	**6.1**	**<.05**	2.5	8.4	1.5	not
ON THE ORDER OF	R	PP	9.3	3.8	0.6	not	10.6	0.0	**6.3**	**<.05**	15.4	1.9	**6.2**	**<.05**	22.2	0.0	**14.0**	**<.001**
ON THE OTHER HAND	T	PP	64.8	139.3	**6.6**	**<.05**	10.6	60.1	**15.3**	**<0.0001**	68.4	167.2	**20.8**	**<0.0001**	138.3	177.0	2.1	not
ON THE OTHER HAND ARE	T	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	2.1	1.2	not
ON THE OTHER HAND IS	T	PP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	7.4	**4.9**	**<.05**	4.9	0.0	3.1	not
ON THE OTHER HAND THE	T	PP	13.9	22.6	0.5	not	5.3	8.9	0.4	not	26.5	40.9	1.5	not	24.7	40.0	1.6	not
ON THE OTHER SIDE	T	PP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	13.2	1.9	**4.9**	**<.05**	14.8	6.3	1.6	not
ON THE OTHER SIDE OF THE	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	8.8	1.9	2.5	not	2.5	0.0	1.6	not
ON THE PART OF	T	PP	0.0	15.1	**4.8**	**<.05**	0.0	15.6	**8.5**	**<.01**	0.0	27.9	**18.3**	**<0.0001**	0.0	33.7	**19.7**	**<0.0001**
ON THE PART OF THE	T	PP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	9.3	**6.1**	**<.05**	0.0	8.4	**4.9**	**<.05**
ON THE PERFORMANCE OF	R	PP	4.6	15.1	1.4	not	0.0	4.5	2.4	not	2.2	7.4	1.4	not	0.0	10.5	**6.2**	**<.05**
ON THE RELATIONSHIP BETWEEN	T	PP	4.6	41.4	**7.8**	**<.01**	2.7	6.7	0.7	not	15.4	31.6	2.8	not	9.9	31.6	**5.2**	**<.05**
ON THE ROLE OF	R	PP	9.3	52.7	**7.8**	**<.01**	0.0	2.2	1.2	not	2.2	7.4	1.4	not	7.4	35.8	**8.7**	**<.01**
ON THE SURFACE OF	R	PP	37.0	0.0	**12.8**	**<.001**	10.6	0.0	**6.3**	**<.05**	15.4	0.0	**11.0**	**<.001**	24.7	0.0	**15.5**	**<0.0001**
ON THE SURFACE OF THE	R	PP	32.4	0.0	**11.2**	**<.001**	5.3	0.0	3.1	not	6.6	0.0	**4.7**	**<.05**	7.4	0.0	**4.7**	**<.05**
ON THE USE OF	R	PP	9.3	22.6	1.4	not	13.3	11.1	0.1	not	0.0	1.9	1.2	not	7.4	31.6	**6.9**	**<.01**
ONE OF THE KEY	P	NP	4.6	11.3	0.7	not	2.7	0.0	1.6	not	2.2	5.6	0.7	not	4.9	10.5	0.9	not
ONE OF THE MAIN	P	NP	4.6	26.4	**3.9**	**<.05**	2.7	6.7	0.7	not	0.0	5.6	3.7	not	12.3	6.3	0.9	not
ONE OF THE MOST	P	NP	138.8	71.5	**5.3**	**<.05**	2.7	35.6	**13.4**	**<.001**	11.0	14.9	0.3	not	24.7	40.0	1.6	not
ONE OF THE MOST IMPORTANT	P	NP	37.0	22.6	0.8	not	0.0	4.5	2.4	not	4.4	1.9	0.5	not	9.9	8.4	0.1	not
ONE OF THE REASONS FOR THE	T	NP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	0.0	1.9	1.2	not	2.5	0.0	1.6	not
ONE OF THE TWO	R	NP	0.0	0.0	0.0	not	18.6	11.1	0.8	not	6.6	7.4	0.0	not	7.4	4.2	0.4	not
ONE STANDARD DEVIATION ABOVE THE MEAN	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	13.0	**8.6**	**<.01**	0.0	0.0	0.0	not
ONE STANDARD DEVIATION BELOW THE MEAN	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	13.0	**8.6**	**<.01**	0.0	0.0	0.0	not
ONE WAY ANALYSIS OF VARIANCE	R	NP	0.0	0.0	0.0	not	24.0	2.2	**8.9**	**<.01**	0.0	5.6	3.7	not	0.0	0.0	0.0	not
OUR FINDINGS SUGGEST THAT	T	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	4.4	5.6	0.1	not	7.4	46.4	**13.5**	**<.001**
OUR RESULTS INDICATE THAT	T	VP	4.6	11.3	0.7	not	0.0	0.0	0.0	not	6.6	9.3	0.2	not	9.9	46.4	**11.0**	**<.001**
OUR RESULTS SHOW THAT	T	VP	0.0	11.3	3.6	not	0.0	2.2	1.2	not	0.0	9.3	**6.1**	**<.05**	17.3	27.4	1.0	not
OUR RESULTS SUGGEST THAT	T	VP	0.0	15.1	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	5.6	3.7	not	34.6	56.9	2.4	not
OUR RESULTS SUGGEST THAT THE	T	VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	4.9	8.4	0.4	not
OUR UNDERSTANDING OF HOW	T	NP	4.6	22.6	3.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	21.1	**12.3**	**<.001**
OUR UNDERSTANDING OF THE	T	NP	9.3	45.2	**6.0**	**<.05**	5.3	4.5	0.0	not	0.0	1.9	1.2	not	4.9	44.3	**15.4**	**<0.0001**
OVER A PERIOD OF	R	PP	4.6	11.3	0.7	not	5.3	11.1	0.8	not	8.8	5.6	0.4	not	9.9	0.0	**6.2**	**<.05**

OVER THE COURSE OF	R	PP	18.5	22.6	0.1	not	21.3	33.4	1.1	not	35.3	27.9	0.4	not	19.8	10.5	1.3	not
OVER THE COURSE OF THE	R	PP	9.3	3.8	0.6	not	13.3	11.1	0.1	not	26.5	11.1	3.2	not	9.9	4.2	1.0	not
OVER THE PAST TWO DECADES	R	PP	4.6	15.1	1.4	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	6.3	3.7	not
PARTICIPANTS WERE ASKED TO	R	VP	0.0	3.8	1.2	not	5.3	42.3	**13.0**	**<.001**	0.0	5.6	3.7	not	0.0	0.0	0.0	not
PARTICIPANTS WERE RANDOMLY ASSIGNED TO	R	VP	0.0	0.0	0.0	not	2.7	24.5	**8.1**	**<.01**	0.0	1.9	1.2	not	0.0	0.0	0.0	not
PER CENT OF THE	R	PP	0.0	18.8	**6.0**	**<.05**	0.0	29.0	**15.8**	**<0.0001**	0.0	72.5	**47.7**	**<0.0001**	0.0	23.2	**13.6**	**<.001**
PLAY A CRITICAL ROLE IN	P	NP_VP	13.9	3.8	1.5	not	0.0	2.2	1.2	not	2.2	0.0	1.6	not	12.3	2.1	3.6	not
PLAY A KEY ROLE	P	NP_VP	13.9	11.3	0.1	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	19.8	6.3	3.2	not
PLAY A KEY ROLE IN	P	NP_VP	13.9	11.3	0.1	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	17.3	6.3	2.3	not
PLAY A KEY ROLE IN THE	P	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	9.9	4.2	1.0	not
PLAY A MORE IMPORTANT ROLE IN	P	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	0.0	5.6	3.7	not	2.5	2.1	0.0	not
PLAY AN IMPORTANT ROLE	P	NP_VP	64.8	26.4	**4.0**	**<.05**	2.7	2.2	0.0	not	11.0	9.3	0.1	not	34.6	27.4	0.4	not
PLAY AN IMPORTANT ROLE IN	P	NP_VP	50.9	22.6	2.7	not	2.7	2.2	0.0	not	11.0	9.3	0.1	not	29.6	27.4	0.0	not
PLAY AN IMPORTANT ROLE IN THE	P	NP_VP	18.5	7.5	1.2	not	0.0	2.2	1.2	not	4.4	1.9	0.5	not	14.8	6.3	1.6	not
PLAYED AN IMPORTANT ROLE IN	P	NP_VP	4.6	3.8	0.0	not	0.0	0.0	0.0	not	2.2	3.7	0.2	not	7.4	4.2	0.4	not
PLAYS A KEY ROLE IN	P	NP_VP	9.3	7.5	0.0	not	0.0	0.0	0.0	not	2.2	0.0	1.6	not	2.5	2.1	0.0	not
PLAYS A SIGNIFICANT ROLE IN	P	NP_VP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	2.1	0.0	not
PLAYS AN IMPORTANT ROLE	P	NP_VP	32.4	15.1	1.6	not	0.0	2.2	1.2	not	13.2	3.7	2.8	not	44.5	8.4	**12.0**	**<.001**
PLAYS AN IMPORTANT ROLE IN	P	NP_VP	27.8	11.3	1.7	not	0.0	2.2	1.2	not	8.8	1.9	2.5	not	34.6	4.2	**12.1**	**<.001**
PLAYS AN IMPORTANT ROLE IN THE	P	NP_VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	12.3	0.0	**7.8**	**<.01**
PLOTTED AS A FUNCTION OF	R	NP_VP	4.6	0.0	1.6	not	8.0	0.0	**4.7**	**<.05**	30.9	0.0	**21.9**	**<0.0001**	9.9	0.0	**6.2**	**<.05**
POINT LIKERT TYPE SCALE	R	NP	0.0	0.0	0.0	not	0.0	69.1	**37.7**	**<0.0001**	0.0	0.0	0.0	not	0.0	2.1	1.2	not
POSITIVELY AND SIGNIFICANTLY RELATED TO	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	14.9	**9.8**	**<.01**	0.0	2.1	1.2	not
PRESENCE OR ABSENCE OF	R	NP	4.6	3.8	0.0	not	13.3	8.9	0.4	not	17.6	16.7	0.0	not	2.5	4.2	0.2	not
PREVIOUS STUDIES HAVE SHOWN THAT	T	VP	13.9	0.0	**4.8**	**<.05**	8.0	0.0	**4.7**	**<.05**	4.4	0.0	3.1	not	9.9	4.2	1.0	not
PURPOSE OF THIS STUDY	T	NP	46.3	37.6	0.2	not	0.0	20.0	**10.9**	**<.001**	4.4	1.9	0.5	not	2.5	14.8	**4.2**	**<.05**
RANDOMLY ASSIGNED TO ONE OF	R	NP_VP	0.0	0.0	0.0	not	0.0	22.3	**12.2**	**<.001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
REFERRED TO AS THE	T	VP	27.8	15.1	0.9	not	8.0	11.1	0.2	not	15.4	1.9	**6.2**	**<.05**	7.4	2.1	1.4	not
REFERS TO THE EXTENT TO WHICH	R	NP_VP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
RESEARCH HAS FOCUSED ON	T	VP	4.6	41.4	**7.8**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
RESEARCH HAS SHOWN THAT	T	VP	4.6	26.4	**3.9**	**<.05**	0.0	6.7	3.6	not	0.0	9.3	**6.1**	**<.05**	0.0	21.1	**12.3**	**<.001**
RESEARCH IN THIS AREA	T	NP	4.6	18.8	2.2	not	0.0	4.5	2.4	not	0.0	0.0	0.0	not	2.5	21.1	**7.2**	**<.01**
RESEARCH IS NEEDED TO	P	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	12.3	46.4	**9.0**	**<.01**
RESPONDENTS WERE ASKED TO	R	VP	0.0	0.0	0.0	not	0.0	51.2	**28.0**	**<0.0001**	0.0	1.9	1.2	not	0.0	0.0	0.0	not
RESULT IS CONSISTENT WITH	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	11.0	7.4	0.3	not	9.9	12.6	0.1	not
RESULTED IN A SIGNIFICANT	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	19.9	18.6	0.0	not	2.5	0.0	1.6	not
RESULTS ARE CONSISTENT WITH	T	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	24.3	24.2	0.0	not	27.2	23.2	0.1	not
RESULTS ARE CONSISTENT WITH THE	T	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	17.6	11.1	0.7	not	9.9	12.6	0.1	not
RESULTS ARE PRESENTED IN TABLE	T	NP_VP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	2.2	24.2	**10.2**	**<.01**	0.0	0.0	0.0	not
RESULTS ARE SHOWN IN TABLE	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	0.0	14.9	**9.8**	**<.01**	0.0	2.1	1.2	not
RESULTS INDICATE THAT THE	T	VP	0.0	7.5	2.4	not	0.0	4.5	2.4	not	24.3	22.3	0.0	not	22.2	21.1	0.0	not
RESULTS OF THIS STUDY	T	NP	9.3	7.5	0.0	not	0.0	2.2	1.2	not	0.0	5.6	3.7	not	37.0	105.4	**14.7**	**<.001**
RESULTS OF THIS STUDY SUGGEST THAT	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	14.8	**4.2**	**<.05**
RESULTS SHOW THAT THE	T	VP	23.1	3.8	3.8	not	2.7	4.5	0.2	not	22.1	13.0	1.2	not	19.8	23.2	0.1	not
RESULTS SHOWED THAT THE	T	VP	37.0	0.0	**12.8**	**<.001**	0.0	4.5	2.4	not	11.0	11.1	0.0	not	14.8	2.1	**4.8**	**<.05**
RESULTS SUGGEST THAT THE	T	VP	0.0	7.5	2.4	not	0.0	4.5	2.4	not	22.1	20.4	0.0	not	29.6	16.9	1.6	not
ROOT MEAN SQUARE ERROR	R	NP	0.0	0.0	0.0	not	0.0	17.8	**9.7**	**<.01**	2.2	18.6	**7.1**	**<.01**	0.0	0.0	0.0	not
ROOT MEAN SQUARE ERROR OF	R	NP	0.0	0.0	0.0	not	0.0	17.8	**9.7**	**<.01**	2.2	18.6	**7.1**	**<.01**	0.0	0.0	0.0	not
RULE OUT THE POSSIBILITY THAT	P	NP_VP	0.0	15.1	**4.8**	**<.05**	0.0	2.2	1.2	not	0.0	11.1	**7.3**	**<.01**	2.5	4.2	0.2	not

Phrase																		
SHED LIGHT ON THE	T	VP	4.6	52.7	**10.9**	**<.001**	2.7	6.7	0.7	not	2.2	3.7	0.2	not	2.5	19.0	**6.2**	**<.05**
SHOULD BE ABLE TO	P	VP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	4.4	1.9	0.5	not	4.9	6.3	0.1	not
SHOULD BE MORE LIKELY TO	P	VP	0.0	7.5	2.4	not	0.0	4.5	2.4	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
SHOULD BE NOTED THAT	P	VP	18.5	7.5	1.2	not	42.6	22.3	2.7	not	83.8	29.7	**13.4**	**<.001**	44.5	23.2	3.0	not
SHOWS THAT THERE IS A	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	5.6	0.1	not	9.9	2.1	2.4	not
SIGNIFICANT DIFFERENCES IN THE	R	NP	0.0	0.0	0.0	not	2.7	17.8	**5.0**	**<.05**	26.5	13.0	2.3	not	2.5	2.1	0.0	not
SIGNIFICANT DIFFERENCES WERE FOUND	R	NP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	11.0	16.7	0.6	not	2.5	0.0	1.6	not
SIGNIFICANT EFFECT ON THE	R	NP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	15.4	11.1	0.3	not	7.4	4.2	0.4	not
SIGNIFICANTLY LOWER THAN THAT OF	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	1.9	**4.9**	**<.05**	7.4	0.0	**4.7**	**<.05**
SIGNIFICANTLY LOWER THAN THAT OF THE	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	11.0	1.9	3.6	not	7.4	0.0	**4.7**	**<.05**
SIMILAR TO THAT OF	T	NP	41.6	3.8	**9.1**	**<.01**	13.3	8.9	0.4	not	52.9	7.4	**19.5**	**<0.0001**	32.1	6.3	**8.4**	**<.01**
SIMILAR TO THAT OF THE	R	NP	23.1	0.0	**8.0**	**<.01**	2.7	2.2	0.0	not	28.7	0.0	**20.3**	**<0.0001**	12.3	0.0	**7.8**	**<.01**
SIMILAR TO THOSE OF	T	NP	4.6	3.8	0.0	not	2.7	4.5	0.2	not	17.6	7.4	2.1	not	17.3	4.2	3.8	not
STANDARD DEVIATION INCREASE IN	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	37.2	**24.4**	**<0.0001**	0.0	0.0	0.0	not
STANDARD DEVIATION OF THE	R	NP	0.0	0.0	0.0	not	8.0	2.2	1.4	not	19.9	14.9	0.4	not	0.0	0.0	0.0	not
STATISTICAL ANALYSIS WAS PERFORMED USING	R	VP	0.0	0.0	0.0	not	24.0	0.0	**14.1**	**<.001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
STUDIES HAVE DEMONSTRATED THAT	T	VP	37.0	3.8	**7.7**	**<.01**	0.0	0.0	0.0	not	6.6	0.0	**4.7**	**<.05**	12.3	2.1	3.6	not
STUDIES HAVE FOCUSED ON	T	VP	23.1	48.9	2.2	not	2.7	0.0	1.6	not	0.0	0.0	0.0	not	4.9	10.5	0.9	not
STUDIES HAVE FOUND THAT	T	VP	18.5	15.1	0.1	not	0.0	4.5	2.4	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
STUDIES HAVE SHOWN THAT	T	VP	101.8	33.9	**8.6**	**<.01**	8.0	2.2	1.4	not	6.6	1.9	1.4	not	27.2	8.4	**4.6**	**<.05**
STUDY CONTRIBUTES TO THE	T	VP	0.0	48.9	**15.5**	**<0.0001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	29.5	**17.3**	**<0.0001**
STUDY IS THE FIRST	P	VP	0.0	18.8	**6.0**	**<.05**	0.0	0.0	0.0	not	0.0	3.7	2.4	not	9.9	16.9	0.8	not
STUDY WAS TO INVESTIGATE THE	T	VP	27.8	0.0	**9.6**	**<.01**	0.0	2.2	1.2	not	0.0	1.9	1.2	not	4.9	2.1	0.5	not
SUCH A WAY THAT	T	PP	9.3	3.8	0.6	not	10.6	0.0	**6.3**	**<.05**	6.6	7.4	0.0	not	12.3	4.2	1.8	not
TAKING INTO ACCOUNT THE	R	VP	13.9	3.8	1.5	not	5.3	8.9	0.4	not	6.6	3.7	0.4	not	19.8	6.3	3.2	not
TEND TO BE MORE	P	VP	0.0	7.5	2.4	not	0.0	8.9	**4.9**	**<.05**	0.0	7.4	**4.9**	**<.05**	4.9	21.1	**4.6**	**<.05**
THAN IN THE CASE OF	T	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	9.3	0.9	not	4.9	2.1	0.5	not
THAN THAT IN THE	R	PP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	30.9	9.3	**6.1**	**<.05**	9.9	2.1	2.4	not
THAN THAT OF THE	R	PP	23.1	0.0	**8.0**	**<.01**	13.3	8.9	0.4	not	86.0	26.0	**17.0**	**<0.0001**	64.2	10.5	**19.1**	**<0.0001**
THAN THOSE IN THE	R	PP	4.6	3.8	0.0	not	2.7	2.2	0.0	not	39.7	16.7	**4.8**	**<.05**	32.1	2.1	**14.2**	**<.001**
THAN THOSE OF THE	R	PP	4.6	3.8	0.0	not	2.7	0.0	1.6	not	41.9	9.3	**11.3**	**<.001**	9.9	8.4	0.1	not
THAT CAN BE USED	R	VP	13.9	18.8	0.2	not	10.6	8.9	0.1	not	0.0	1.9	1.2	not	4.9	4.2	0.0	not
THAT CAN BE USED TO	R	VP	13.9	11.3	0.1	not	5.3	4.5	0.0	not	0.0	1.9	1.2	not	0.0	2.1	1.2	not
THAT HIGHER LEVELS OF	R	NP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	13.0	**8.6**	**<.01**	2.5	4.2	0.2	not
THAT IN ADDITION TO	T	PP	13.9	7.5	0.5	not	0.0	2.2	1.2	not	4.4	1.9	0.5	not	4.9	14.8	2.2	not
THAT IN THE ABSENCE OF	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	6.6	3.7	0.4	not	12.3	4.2	1.8	not
THAT IT IS NOT	T	VP	9.3	3.8	0.6	not	10.6	2.2	2.5	not	4.4	7.4	0.4	not	2.5	19.0	**6.2**	**<.05**
THAT MOST OF THE	T	NP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	8.8	14.9	0.8	not	14.8	6.3	1.6	not
THAT OF THE WILD TYPE	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	24.3	0.0	**17.2**	**<0.0001**	0.0	0.0	0.0	not
THAT THE AMOUNT OF	R	NP	4.6	7.5	0.2	not	0.0	2.2	1.2	not	13.2	9.3	0.3	not	9.9	4.2	1.0	not
THAT THE EFFECT OF	R	NP	0.0	11.3	3.6	not	0.0	6.7	3.6	not	6.6	39.0	**12.3**	**<.001**	17.3	23.2	0.4	not
THAT THE EFFECTS OF	R	NP	4.6	33.9	**5.8**	**<.05**	2.7	0.0	1.6	not	4.4	13.0	2.2	not	14.8	21.1	0.5	not
THAT THE IMPACT OF	R	NP	0.0	15.1	**4.8**	**<.05**	0.0	4.5	2.4	not	2.2	13.0	**4.1**	**<.05**	0.0	12.6	**7.4**	**<.01**
THAT THE LEVEL OF	R	NP	4.6	7.5	0.2	not	0.0	4.5	2.4	not	6.6	14.9	1.6	not	4.9	14.8	2.2	not
THAT THE NUMBER OF	R	NP	4.6	15.1	1.4	not	5.3	15.6	2.1	not	11.0	9.3	0.1	not	2.5	21.1	**7.2**	**<.01**
THAT THE POSITIVE RELATIONSHIP BETWEEN	T	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	13.0	**8.6**	**<.01**	0.0	2.1	1.2	not
THAT THE PRESENCE OF	R	NP	4.6	11.3	0.7	not	2.7	0.0	1.6	not	11.0	9.3	0.1	not	9.9	8.4	0.1	not
THAT THE RELATIONSHIP BETWEEN	T	NP	0.0	15.1	**4.8**	**<.05**	0.0	4.5	2.4	not	2.2	14.9	**5.1**	**<.05**	9.9	16.9	0.8	not
THAT THE USE OF	R	NP	9.3	26.4	2.0	not	2.7	4.5	0.2	not	2.2	3.7	0.2	not	17.3	12.6	0.3	not

Phrase																		
THAT THERE IS A	P	VP	41.6	30.1	0.4	not	2.7	22.3	**7.0**	**<.01**	44.1	52.0	0.3	not	49.4	25.3	3.5	not
THAT THERE IS NO	P	VP	9.3	7.5	0.0	not	0.0	4.5	2.4	not	11.0	27.9	3.7	not	0.0	19.0	**11.1**	**<.001**
THAT THIS IS NOT THE CASE	T	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	5.6	3.7	not	0.0	8.4	**4.9**	**<.05**
THE ABILITY OF THE	R	NP	18.5	7.5	1.2	not	8.0	2.2	1.4	not	4.4	1.9	0.5	not	24.7	8.4	3.7	not
THE ABSENCE OF ANY	R	NP	9.3	3.8	0.6	not	0.0	2.2	1.2	not	11.0	3.7	1.9	not	4.9	8.4	0.4	not
THE ACCURACY OF THE	R	NP	4.6	7.5	0.2	not	10.6	13.4	0.1	not	11.0	1.9	3.6	not	17.3	6.3	2.3	not
THE ADDITION OF THE	R	NP	4.6	0.0	1.6	not	10.6	0.0	**6.3**	**<.05**	4.4	16.7	3.7	not	12.3	0.0	**7.8**	**<.01**
THE AIM OF THIS	T	NP	92.5	45.2	**4.0**	**<.05**	2.7	4.5	0.2	not	0.0	1.9	1.2	not	0.0	10.5	**6.2**	**<.05**
THE AIM OF THIS STUDY	T	NP	69.4	7.5	**14.1**	**<.001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	6.3	3.7	not
THE AIM OF THIS STUDY WAS TO	T	NP_VP	50.9	3.8	**11.9**	**<.001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
THE AMOUNT OF TIME	R	NP	4.6	15.1	1.4	not	5.3	11.1	0.8	not	4.4	1.9	0.5	not	2.5	2.1	0.0	not
THE AMPLITUDE OF THE	R	NP	9.3	0.0	3.2	not	8.0	0.0	**4.7**	**<.05**	44.1	1.9	**24.5**	**<0.0001**	19.8	0.0	**12.4**	**<.001**
THE ANALYSIS OF THE	R	NP	13.9	18.8	0.2	not	21.3	13.4	0.8	not	15.4	40.9	**5.8**	**<.05**	22.2	16.9	0.3	not
THE APPLICATION OF THE	R	NP	9.3	7.5	0.0	not	10.6	2.2	2.5	not	4.4	0.0	3.1	not	9.9	4.2	1.0	not
THE AVERAGE NUMBER OF	R	NP	0.0	0.0	0.0	not	2.7	46.8	**19.0**	**<0.0001**	2.2	27.9	**12.4**	**<.001**	2.5	6.3	0.8	not
THE AVERAGE VARIANCE EXTRACTED	R	NP	0.0	0.0	0.0	not	0.0	17.8	**9.7**	**<.01**	0.0	26.0	**17.1**	**<0.0001**	0.0	0.0	0.0	not
THE BASE OF THE	R	NP	4.6	0.0	1.6	not	26.6	2.2	**10.2**	**<.01**	26.5	0.0	**18.8**	**<0.0001**	9.9	0.0	**6.2**	**<.05**
THE BASIS OF THE	T	NP	46.3	18.8	2.9	not	31.9	31.2	0.0	not	22.1	29.7	0.6	not	22.2	29.5	0.4	not
THE BASIS OF THEIR	T	NP	9.3	0.0	3.2	not	2.7	8.9	1.4	not	8.8	3.7	1.1	not	4.9	8.4	0.4	not
THE BASIS OF THESE	T	NP	4.6	7.5	0.2	not	2.7	2.2	0.0	not	2.2	7.4	1.4	not	4.9	2.1	0.5	not
THE BEGINNING OF THE	R	NP	0.0	11.3	3.6	not	18.6	35.6	2.2	not	13.2	26.0	2.1	not	12.3	31.6	3.8	not
THE BOTTOM OF THE	R	NP	0.0	3.8	1.2	not	34.6	2.2	**14.4**	**<.001**	13.2	13.0	0.0	not	12.3	6.3	0.9	not
THE CASE OF THE	T	NP	4.6	11.3	0.7	not	18.6	4.5	**3.9**	**<.05**	39.7	5.6	**14.6**	**<.001**	44.5	25.3	2.3	not
THE CENTER OF THE	R	NP	0.0	0.0	0.0	not	66.5	2.2	**32.0**	**<0.0001**	35.3	1.9	**18.7**	**<0.0001**	27.2	0.0	**17.1**	**<0.0001**
THE CENTRE OF THE	R	NP	0.0	0.0	0.0	not	24.0	0.0	**14.1**	**<.001**	44.1	5.6	**17.2**	**<0.0001**	22.2	0.0	**14.0**	**<.001**
THE CHANGE IN THE	R	NP	32.4	0.0	**11.2**	**<.001**	0.0	15.6	**8.5**	**<.01**	17.6	13.0	0.3	not	24.7	6.3	**5.2**	**<.05**
THE CHANGES IN THE	R	NP	18.5	3.8	2.6	not	0.0	2.2	1.2	not	8.8	0.0	**6.3**	**<.05**	22.2	6.3	**4.2**	**<.05**
THE CHARACTERISTICS OF THE	R	NP	9.3	26.4	2.0	not	5.3	24.5	**5.4**	**<.05**	15.4	9.3	0.8	not	14.8	21.1	0.5	not
THE COMPLEXITY OF THE	R	NP	4.6	7.5	0.2	not	0.0	8.9	**4.9**	**<.05**	4.4	5.6	0.1	not	7.4	16.9	1.6	not
THE COMPOSITION OF THE	R	NP	4.6	7.5	0.2	not	16.0	2.2	**4.9**	**<.05**	11.0	9.3	0.1	not	14.8	8.4	0.8	not
THE CONDITIONS UNDER WHICH	R	NP	0.0	22.6	**7.1**	**<.01**	0.0	0.0	0.0	not	2.2	1.9	0.0	not	4.9	19.0	3.8	not
THE CONTENT OF THE	R	NP	0.0	3.8	1.2	not	0.0	11.1	**6.1**	**<.05**	2.2	18.6	**7.1**	**<.01**	0.0	16.9	**9.9**	**<.01**
THE CONTEXT IN WHICH	T	NP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	0.0	3.7	2.4	not	0.0	10.5	**6.2**	**<.05**
THE CONTEXT OF A	T	NP	4.6	22.6	3.0	not	0.0	11.1	**6.1**	**<.05**	2.2	3.7	0.2	not	0.0	12.6	**7.4**	**<.01**
THE CONTEXT OF THE	T	NP	4.6	33.9	**5.8**	**<.05**	0.0	17.8	**9.7**	**<.01**	4.4	9.3	0.9	not	17.3	35.8	2.9	not
THE CONTRIBUTION OF THE	T	NP	4.6	11.3	0.7	not	5.3	0.0	3.1	not	11.0	13.0	0.1	not	7.4	12.6	0.6	not
THE CORRELATION BETWEEN THE	R	NP	4.6	0.0	1.6	not	13.3	11.1	0.1	not	15.4	13.0	0.1	not	4.9	0.0	3.1	not
THE COURSE OF THE	R	NP	9.3	7.5	0.0	not	24.0	13.4	1.3	not	39.7	18.6	**3.9**	**<.05**	17.3	8.4	1.4	not
THE COURSE OF THE STUDY	R	NP	0.0	0.0	0.0	not	10.6	0.0	**6.3**	**<.05**	4.4	3.7	0.0	not	4.9	0.0	3.1	not
THE CRYSTAL STRUCTURE OF THE	R	NP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	4.4	0.0	3.1	not	9.9	0.0	**6.2**	**<.05**
THE DATA FROM THE	T	NP	0.0	0.0	0.0	not	16.0	8.9	0.8	not	4.4	5.6	0.1	not	2.5	10.5	2.3	not
THE DATA WERE COLLECTED	R	VP	0.0	0.0	0.0	not	2.7	33.4	**12.3**	**<.001**	0.0	3.7	2.4	not	0.0	4.2	2.5	not
THE DEGREE TO WHICH	R	NP	9.3	52.7	**7.8**	**<.01**	5.3	53.5	**18.2**	**<0.0001**	0.0	29.7	**19.6**	**<0.0001**	9.9	35.8	**6.7**	**<.01**
THE DEGREE TO WHICH A	R	NP	4.6	7.5	0.2	not	0.0	6.7	3.6	not	0.0	5.6	3.7	not	2.5	4.2	0.2	not
THE DEGREE TO WHICH THE	R	NP	0.0	7.5	2.4	not	2.7	13.4	3.1	not	0.0	11.1	**7.3**	**<.01**	2.5	14.8	**4.2**	**<.05**
THE DEPENDENT VARIABLE IN	R	NP	0.0	0.0	0.0	not	0.0	33.4	**18.2**	**<0.0001**	0.0	18.6	**12.2**	**<.001**	0.0	2.1	1.2	not
THE DEPENDENT VARIABLE IS	R	NP	0.0	0.0	0.0	not	0.0	35.6	**19.5**	**<0.0001**	2.2	20.4	**8.1**	**<.01**	0.0	0.0	0.0	not
THE DEPENDENT VARIABLE IS THE	R	NP	0.0	0.0	0.0	not	0.0	15.6	**8.5**	**<.01**	2.2	9.3	2.3	not	0.0	0.0	0.0	not
THE DETAILS OF THE	T	NP	0.0	3.8	1.2	not	10.6	8.9	0.1	not	15.4	3.7	**3.9**	**<.05**	4.9	6.3	0.1	not

THE DEVELOPMENT OF A	R	NP	23.1	30.1	0.2	not	2.7	4.5	0.2	not	2.2	1.9	0.0	not	17.3	12.6	0.3	not
THE DEVELOPMENT OF THE	R	NP	13.9	11.3	0.1	not	2.7	15.6	**4.1**	**<.05**	2.2	5.6	0.7	not	4.9	25.3	**6.4**	**<.05**
THE DIAMETER OF THE	R	NP	0.0	0.0	0.0	not	24.0	0.0	**14.1**	**<.001**	15.4	0.0	**11.0**	**<.001**	17.3	0.0	**10.9**	**<.001**
THE DIFFERENCE BETWEEN THE	R	NP	13.9	0.0	**4.8**	**<.05**	26.6	17.8	0.7	not	41.9	35.3	0.3	not	32.1	4.2	**10.9**	**<.001**
THE DIFFERENCE IN THE	R	NP	9.3	0.0	3.2	not	24.0	8.9	3.0	not	15.4	16.7	0.0	not	27.2	6.3	**6.2**	**<.05**
THE DIFFERENCES BETWEEN THE	R	NP	9.3	7.5	0.0	not	18.6	6.7	2.4	not	35.3	22.3	1.5	not	12.3	8.4	0.3	not
THE DIFFERENCES IN THE	R	NP	4.6	0.0	1.6	not	2.7	11.1	2.2	not	22.1	13.0	1.2	not	9.9	6.3	0.3	not
THE DIRECTION OF THE	R	NP	4.6	15.1	1.4	not	2.7	2.2	0.0	not	13.2	5.6	1.6	not	7.4	8.4	0.0	not
THE DISTANCE BETWEEN THE	R	NP	0.0	0.0	0.0	not	24.0	8.9	3.0	not	15.4	1.9	**6.2**	**<.05**	7.4	0.0	**4.7**	**<.05**
THE DISTRIBUTION OF THE	R	NP	4.6	0.0	1.6	not	16.0	20.0	0.2	not	33.1	11.1	**5.7**	**<.05**	7.4	0.0	**4.7**	**<.05**
THE DURATION OF THE	R	NP	4.6	3.8	0.0	not	21.3	13.4	0.8	not	22.1	20.4	0.0	not	12.3	6.3	0.9	not
THE DYNAMICS OF THE	R	NP	18.5	3.8	2.6	not	5.3	0.0	3.1	not	13.2	0.0	**9.4**	**<.01**	4.9	6.3	0.1	not
THE EARLY STAGES OF	R	NP	4.6	3.8	0.0	not	8.0	4.5	0.4	not	6.6	1.9	1.4	not	14.8	10.5	0.3	not
THE EDGE OF THE	R	NP	4.6	3.8	0.0	not	13.3	0.0	**7.9**	**<.01**	30.9	1.9	**15.8**	**<0.0001**	4.9	0.0	3.1	not
THE EFFECT OF A	R	NP	41.6	7.5	**6.4**	**<.05**	2.7	13.4	3.1	not	22.1	37.2	1.9	not	14.8	6.3	1.6	not
THE EFFECT OF THE	R	NP	55.5	26.4	2.6	not	31.9	22.3	0.7	not	52.9	50.2	0.0	not	39.5	16.9	**4.1**	**<.05**
THE EFFECTIVENESS OF THE	R	NP	0.0	11.3	3.6	not	0.0	6.7	3.6	not	0.0	9.3	**6.1**	**<.05**	2.5	16.9	**5.1**	**<.05**
THE EFFECTS OF DIFFERENT	R	NP	4.6	15.1	1.4	not	8.0	2.2	1.4	not	2.2	3.7	0.2	not	2.5	14.8	**4.2**	**<.05**
THE EFFECTS OF THE	R	NP	55.5	33.9	1.3	not	21.3	17.8	0.1	not	26.5	61.3	**6.9**	**<.01**	32.1	23.2	0.6	not
THE EFFECTS OF THESE	R	NP	9.3	15.1	0.3	not	2.7	6.7	0.7	not	0.0	7.4	**4.9**	**<.05**	0.0	4.2	2.5	not
THE END OF THE	R	NP	27.8	11.3	1.7	not	63.9	80.2	0.8	not	70.6	39.0	**4.6**	**<.05**	27.2	29.5	0.0	not
THE END OF THE SEMESTER	R	NP	0.0	0.0	0.0	not	0.0	11.1	**6.1**	**<.05**	0.0	3.7	2.4	not	0.0	6.3	3.7	not
THE EVOLUTION OF THE	R	NP	0.0	7.5	2.4	not	16.0	0.0	**9.4**	**<.01**	28.7	0.0	**20.3**	**<0.0001**	29.6	4.2	**9.6**	**<.01**
THE EXISTENCE OF A	R	NP	9.3	7.5	0.0	not	2.7	15.6	**4.1**	**<.05**	6.6	9.3	0.2	not	12.3	12.6	0.0	not
THE EXPLANATORY POWER OF	R	NP	0.0	15.1	**4.8**	**<.05**	0.0	4.5	2.4	not	0.0	14.9	**9.8**	**<.01**	0.0	4.2	2.5	not
THE EXTENT OF THE	T	NP	13.9	7.5	0.5	not	5.3	6.7	0.1	not	6.6	3.7	0.4	not	9.9	6.3	0.3	not
THE EXTENT TO WHICH	R	NP	23.1	195.7	**36.1**	**<0.0001**	5.3	162.6	**73.5**	**<0.0001**	2.2	72.5	**39.9**	**<0.0001**	29.6	97.0	**16.2**	**<0.0001**
THE EXTENT TO WHICH A	R	NP	4.6	11.3	0.7	not	0.0	24.5	**13.4**	**<.001**	0.0	11.1	**7.3**	**<.01**	2.5	8.4	1.5	not
THE EXTENT TO WHICH THE	R	NP	0.0	60.2	**19.1**	**<0.0001**	2.7	55.7	**23.5**	**<0.0001**	0.0	22.3	**14.7**	**<.001**	9.9	14.8	0.4	not
THE EXTENT TO WHICH THEY	R	NP	0.0	15.1	**4.8**	**<.05**	0.0	22.3	**12.2**	**<.001**	0.0	3.7	2.4	not	0.0	6.3	3.7	not
THE FACT THAT IN	P	NP	4.6	3.8	0.0	not	2.7	0.0	1.6	not	8.8	11.1	0.1	not	7.4	12.6	0.6	not
THE FACT THAT THE	P	NP	23.1	33.9	0.5	not	10.6	35.6	**5.7**	**<.05**	64.0	78.0	0.7	not	84.0	67.4	0.8	not
THE FACT THAT THEY	P	NP	4.6	0.0	1.6	not	0.0	8.9	**4.9**	**<.05**	0.0	20.4	**13.4**	**<.001**	2.5	8.4	1.5	not
THE FINDINGS OF PREVIOUS STUDIES	T	NP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	2.5	8.4	1.5	not
THE FINDINGS OF THE	T	NP	0.0	18.8	**6.0**	**<.05**	0.0	4.5	2.4	not	0.0	20.4	**13.4**	**<.001**	12.3	61.1	**15.1**	**<0.0001**
THE FINDINGS OF THE PRESENT STUDY	T	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	21.1	**7.2**	**<.01**
THE FINDINGS OF THIS	T	NP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	2.5	111.7	**57.0**	**<0.0001**
THE FINDINGS OF THIS STUDY	T	NP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	0.0	86.4	**50.6**	**<0.0001**
THE FINDINGS SUGGEST THAT	T	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	25.3	**14.8**	**<.001**
THE FIRST AND SECOND	R	NP	0.0	7.5	2.4	not	13.3	17.8	0.3	not	2.2	9.3	2.3	not	4.9	0.0	3.1	not
THE FIRST AND THE	R	NP	18.5	0.0	**6.4**	**<.05**	13.3	4.5	1.9	not	13.2	3.7	2.8	not	9.9	0.0	**6.2**	**<.05**
THE FIRST AND THE SECOND	R	NP	18.5	0.0	**6.4**	**<.05**	13.3	0.0	**7.9**	**<.01**	13.2	1.9	**4.9**	**<.05**	9.9	0.0	**6.2**	**<.05**
THE FIRST PART OF THE	T	NP	0.0	11.3	3.6	not	2.7	13.4	3.1	not	2.2	3.7	0.2	not	2.5	0.0	1.6	not
THE FOCUS OF THE	T	NP	4.6	3.8	0.0	not	0.0	22.3	**12.2**	**<.001**	6.6	3.7	0.4	not	0.0	0.0	0.0	not
THE FOCUS OF THIS	T	NP	9.3	15.1	0.3	not	5.3	8.9	0.4	not	4.4	0.0	3.1	not	0.0	6.3	3.7	not
THE FOLLOWING CONCLUSIONS CAN BE DRAWN	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	17.3	0.0	**10.9**	**<.001**
THE FOLLOWING HYPOTHESIS IS	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
THE FOLLOWING RESEARCH QUESTIONS	T	NP	0.0	41.4	**13.1**	**<.001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
THE FORM OF A	R	NP	13.9	7.5	0.5	not	2.7	13.4	3.1	not	2.2	16.7	**6.1**	**<.05**	2.5	4.2	0.2	not

Phrase																		
THE FORM OF THE	R	NP	4.6	0.0	1.6	not	13.3	4.5	1.9	not	8.8	3.7	1.1	not	4.9	4.2	0.0	not
THE FORMATION OF A	R	NP	50.9	0.0	**17.6**	**<0.0001**	5.3	4.5	0.0	not	33.1	0.0	**23.5**	**<0.0001**	61.7	0.0	**38.8**	**<0.0001**
THE FORMATION OF THE	R	NP	18.5	0.0	**6.4**	**<.05**	5.3	4.5	0.0	not	17.6	0.0	**12.5**	**<.001**	37.0	4.2	**13.4**	**<.001**
THE GROWTH OF THE	R	NP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	11.0	0.0	**7.8**	**<.01**	9.9	6.3	0.3	not
THE IMPACT OF THE	R	NP	46.3	41.4	0.1	not	8.0	22.3	2.8	not	6.6	27.9	**6.8**	**<.01**	22.2	31.6	0.7	not
THE IMPORTANCE OF THE	P	NP	23.1	30.1	0.2	not	2.7	17.8	**5.0**	**<.05**	2.2	31.6	**14.6**	**<.001**	17.3	37.9	3.4	not
THE IMPORTANT ROLE OF	R	NP	9.3	15.1	0.3	not	0.0	0.0	0.0	not	4.4	3.7	0.0	not	9.9	8.4	0.1	not
THE INCREASE IN THE	R	NP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	15.4	5.6	2.4	not	12.3	2.1	3.6	not
THE INDIRECT EFFECT OF	R	NP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	18.6	**12.2**	**<.001**	0.0	6.3	3.7	not
THE INFLUENCE OF THE	R	NP	69.4	22.6	**6.1**	**<.05**	24.0	11.1	2.0	not	37.5	7.4	**11.0**	**<.001**	32.1	14.8	2.9	not
THE INTENSITY OF THE	R	NP	9.3	0.0	3.2	not	13.3	2.2	3.7	not	33.1	0.0	**23.5**	**<0.0001**	12.3	4.2	1.8	not
THE INTERACTION BETWEEN THE	T	NP	13.9	3.8	1.5	not	2.7	2.2	0.0	not	8.8	9.3	0.0	not	27.2	2.1	**11.4**	**<.001**
THE INTERPRETATION OF THE	T	NP	4.6	0.0	1.6	not	2.7	6.7	0.7	not	6.6	22.3	**4.3**	**<.05**	7.4	0.0	**4.7**	**<.05**
THE INTERPRETATION OF THE RESULTS	T	NP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	4.4	9.3	0.9	not	0.0	0.0	0.0	not
THE LACK OF A	R	NP	4.6	11.3	0.7	not	8.0	4.5	0.4	not	11.0	11.1	0.0	not	51.9	29.5	2.7	not
THE LEADING EDGE OF THE	R	NP	9.3	0.0	3.2	not	10.6	0.0	**6.3**	**<.05**	6.6	0.0	**4.7**	**<.05**	2.5	0.0	1.6	not
THE LENGTH OF THE	R	NP	18.5	3.8	2.6	not	50.6	22.3	**4.7**	**<.05**	30.9	5.6	**9.7**	**<.01**	17.3	10.5	0.7	not
THE LIKELIHOOD OF A	R	NP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	0.0	26.0	**17.1**	**<0.0001**	2.5	14.8	**4.2**	**<.05**
THE LITERATURE ON THE	T	NP	0.0	18.8	**6.0**	**<.05**	0.0	0.0	0.0	not	2.2	1.9	0.0	not	0.0	14.8	**8.6**	**<.01**
THE LOCATION OF THE	R	NP	18.5	0.0	**6.4**	**<.05**	10.6	11.1	0.0	not	13.2	0.0	**9.4**	**<.01**	7.4	6.3	0.0	not
THE MAGNITUDE OF THE	R	NP	18.5	7.5	1.2	not	0.0	8.9	**4.9**	**<.05**	59.6	31.6	**4.3**	**<.05**	54.3	14.8	**10.7**	**<.01**
THE MAIN EFFECT OF	R	NP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	0.0	31.6	**20.8**	**<0.0001**	0.0	4.2	2.5	not
THE MAIN EFFECTS OF	R	NP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	0.0	26.0	**17.1**	**<0.0001**	0.0	6.3	3.7	not
THE MAJORITY OF THE	R	NP	23.1	15.1	0.4	not	10.6	33.4	**5.0**	**<.05**	30.9	18.6	1.5	not	22.2	14.8	0.7	not
THE MECHANICAL BEHAVIOR OF THE	R	NP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	8.8	0.0	**6.3**	**<.05**	9.9	0.0	**6.2**	**<.05**
THE MECHANICAL PROPERTIES OF	R	NP	46.3	0.0	**16.0**	**<0.0001**	16.0	0.0	**9.4**	**<.01**	15.4	0.0	**11.0**	**<.001**	34.6	0.0	**21.7**	**<0.0001**
THE MECHANICAL PROPERTIES OF THE	R	NP	13.9	0.0	**4.8**	**<.05**	16.0	0.0	**9.4**	**<.01**	11.0	0.0	**7.8**	**<.01**	19.8	0.0	**12.4**	**<.001**
THE MIDDLE OF THE	R	NP	0.0	0.0	0.0	not	21.3	8.9	2.2	not	24.3	5.6	**6.3**	**<.05**	12.3	2.1	3.6	not
THE MODERATING EFFECT OF	R	NP	0.0	26.4	**8.3**	**<.01**	0.0	6.7	3.6	not	0.0	31.6	**20.8**	**<0.0001**	0.0	14.8	**8.6**	**<.01**
THE MODERATING EFFECTS OF	R	NP	0.0	26.4	**8.3**	**<.01**	0.0	2.2	1.2	not	0.0	11.1	**7.3**	**<.01**	0.0	12.6	**7.4**	**<.01**
THE MODERATING ROLE OF	R	NP	0.0	22.6	**7.1**	**<.01**	0.0	4.5	2.4	not	0.0	3.7	2.4	not	0.0	12.6	**7.4**	**<.01**
THE MOST LIKELY TO	P	NP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	16.7	**11.0**	**<.001**	2.5	8.4	1.5	not
THE NATURAL LOGARITHM OF THE	R	NP	0.0	0.0	0.0	not	2.7	11.1	2.2	not	4.4	3.7	0.0	not	0.0	0.0	0.0	not
THE NATURE OF THE	R	NP	18.5	67.8	**7.0**	**<.01**	5.3	26.7	**6.3**	**<.05**	13.2	35.3	**5.1**	**<.05**	19.8	65.3	**11.1**	**<.001**
THE NEGATIVE EFFECT OF	R	NP	4.6	15.1	1.4	not	0.0	0.0	0.0	not	0.0	31.6	**20.8**	**<0.0001**	2.5	14.8	**4.2**	**<.05**
THE NEGATIVE EFFECTS OF	R	NP	4.6	15.1	1.4	not	0.0	2.2	1.2	not	0.0	5.6	3.7	not	2.5	19.0	**6.2**	**<.05**
THE NEGATIVE RELATIONSHIP BETWEEN	R	NP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	14.9	**9.8**	**<.01**	2.5	10.5	2.3	not
THE NUMBER OF PARTIES	R	NP	0.0	7.5	2.4	not	0.0	4.5	2.4	not	0.0	14.9	**9.8**	**<.01**	0.0	0.0	0.0	not
THE NUMBER OF YEARS	R	NP	0.0	0.0	0.0	not	2.7	44.6	**17.9**	**<0.0001**	2.2	7.4	1.4	not	0.0	4.2	2.5	not
THE ONE HAND AND	T	PP	0.0	30.1	**9.5**	**<.01**	0.0	17.8	**9.7**	**<.01**	0.0	9.3	**6.1**	**<.05**	2.5	8.4	1.5	not
THE ONSET OF THE	R	NP	0.0	11.3	3.6	not	2.7	15.6	**4.1**	**<.05**	6.6	3.7	0.4	not	2.5	6.3	0.8	not
THE OTHER HAND THE	T	NP	13.9	22.6	0.5	not	5.3	8.9	0.4	not	26.5	40.9	1.5	not	24.7	40.0	1.6	not
THE PART OF THE	T	NP	0.0	7.5	2.4	not	2.7	2.2	0.0	not	0.0	9.3	**6.1**	**<.05**	0.0	8.4	**4.9**	**<.05**
THE PAST TWO DECADES	R	NP	13.9	26.4	0.9	not	0.0	2.2	1.2	not	0.0	11.1	**7.3**	**<.01**	0.0	6.3	3.7	not
THE PERFORMANCE OF THE	R	NP	9.3	7.5	0.0	not	10.6	13.4	0.1	not	8.8	16.7	1.2	not	12.3	27.4	2.5	not
THE POSITION OF THE	R	NP	4.6	0.0	1.6	not	29.3	6.7	**6.4**	**<.05**	35.3	0.0	**25.0**	**<0.0001**	29.6	4.2	**9.6**	**<.01**
THE POSITIVE EFFECT OF	R	NP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	0.0	20.4	**13.4**	**<.001**	0.0	10.5	**6.2**	**<.05**
THE POSITIVE RELATIONSHIP BETWEEN	T	NP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	2.2	31.6	**14.6**	**<.001**	0.0	31.6	**18.5**	**<0.0001**
THE POSSIBILITY THAT THE	P	NP	4.6	15.1	1.4	not	0.0	6.7	3.6	not	13.2	13.0	0.0	not	14.8	16.9	0.1	not

THE PRESENCE OF A	R	NP	55.5	37.6	0.8	not	50.6	24.5	3.8	not	70.6	27.9	**9.6**	**<.01**	71.6	21.1	**12.9**	**<.001**
THE PRESENCE OF AN	R	NP	18.5	7.5	1.2	not	10.6	0.0	**6.3**	**<.05**	8.8	9.3	0.0	not	22.2	4.2	**6.0**	**<.05**
THE PRESENCE OF THE	R	NP	32.4	7.5	**4.1**	**<.05**	8.0	4.5	0.4	not	46.3	16.7	**7.2**	**<.01**	59.3	6.3	**22.1**	**<0.0001**
THE PRESENCE OR ABSENCE OF	R	NP	4.6	3.8	0.0	not	13.3	6.7	0.9	not	17.6	14.9	0.1	not	2.5	4.2	0.2	not
THE PROBABILITY OF BEING	R	NP	0.0	11.3	3.6	not	0.0	11.1	**6.1**	**<.05**	0.0	27.9	**18.3**	**<0.0001**	0.0	2.1	1.2	not
THE PROCESS BY WHICH	R	NP	9.3	15.1	0.3	not	0.0	4.5	2.4	not	2.2	0.0	1.6	not	2.5	2.1	0.0	not
THE PURPOSE OF THIS	T	NP	50.9	41.4	0.2	not	5.3	33.4	**9.1**	**<.01**	2.2	9.3	2.3	not	4.9	19.0	3.8	not
THE PURPOSE OF THIS PAPER IS TO	T	NP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
THE PURPOSE OF THIS STUDY IS TO	T	NP	18.5	22.6	0.1	not	0.0	8.9	**4.9**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
THE PURPOSE OF THIS STUDY WAS TO	T	NP	13.9	3.8	1.5	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	6.3	0.8	not
THE QUALITY OF THE	R	NP	13.9	18.8	0.2	not	13.3	15.6	0.1	not	15.4	11.1	0.3	not	14.8	10.5	0.3	not
THE QUESTION OF HOW	T	NP	0.0	30.1	**9.5**	**<.01**	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	8.4	**4.9**	**<.05**
THE QUESTION OF WHETHER	T	NP	4.6	18.8	2.2	not	0.0	2.2	1.2	not	0.0	14.9	**9.8**	**<.01**	4.9	10.5	0.9	not
THE RATIO OF THE	R	NP	9.3	0.0	3.2	not	24.0	13.4	1.3	not	26.5	7.4	**5.7**	**<.05**	14.8	0.0	**9.3**	**<.01**
THE REASON FOR THIS	T	NP	9.3	0.0	3.2	not	0.0	4.5	2.4	not	15.4	9.3	0.8	not	9.9	4.2	1.0	not
THE RELATIONSHIP BETWEEN THE	T	NP	27.8	22.6	0.1	not	16.0	24.5	0.7	not	26.5	27.9	0.0	not	22.2	21.1	0.0	not
THE RELATIVE ABUNDANCE OF	R	NP	0.0	0.0	0.0	not	13.3	0.0	**7.9**	**<.01**	28.7	0.0	**20.3**	**<0.0001**	17.3	0.0	**10.9**	**<.001**
THE RELATIVE CONTRIBUTIONS OF	R	NP	4.6	0.0	1.6	not	5.3	0.0	3.1	not	8.8	1.9	2.5	not	22.2	2.1	**8.7**	**<.01**
THE RELATIVE IMPORTANCE OF	P	NP	32.4	11.3	2.6	not	2.7	15.6	**4.1**	**<.05**	2.2	13.0	**4.1**	**<.05**	9.9	12.6	0.1	not
THE RELIABILITY OF THE	R	NP	4.6	3.8	0.0	not	2.7	37.9	**14.5**	**<.001**	0.0	3.7	2.4	not	4.9	8.4	0.4	not
THE REMAINDER OF THIS	T	NP	9.3	33.9	3.5	not	0.0	2.2	1.2	not	4.4	9.3	0.9	not	0.0	2.1	1.2	not
THE RESPONSE OF THE	R	NP	27.8	0.0	**9.6**	**<.01**	13.3	0.0	**7.9**	**<.01**	13.2	1.9	**4.9**	**<.05**	24.7	0.0	**15.5**	**<0.0001**
THE REST OF THE	T/R	NP	4.6	26.4	**3.9**	**<.05**	10.6	31.2	**4.2**	**<.05**	24.3	18.6	0.4	not	7.4	14.8	1.1	not
THE RESULT OF A	T	NP	4.6	7.5	0.2	not	0.0	8.9	**4.9**	**<.05**	15.4	13.0	0.1	not	2.5	6.3	0.8	not
THE RESULT OF THE	T	NP	4.6	11.3	0.7	not	2.7	4.5	0.2	not	11.0	7.4	0.3	not	12.3	12.6	0.0	not
THE RESULTS ARE PRESENTED IN	T	NP_VP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	8.8	7.4	0.1	not	0.0	0.0	0.0	not
THE RESULTS ARE SHOWN IN	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	8.8	9.3	0.0	not	4.9	2.1	0.5	not
THE RESULTS FOR THE	T	NP	4.6	0.0	1.6	not	2.7	8.9	1.4	not	15.4	29.7	2.2	not	2.5	4.2	0.2	not
THE RESULTS FROM THE	T	NP	0.0	15.1	**4.8**	**<.05**	5.3	2.2	0.5	not	24.3	20.4	0.2	not	4.9	6.3	0.1	not
THE RESULTS IN TABLE	T	NP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	2.2	39.0	**19.1**	**<0.0001**	9.9	2.1	2.4	not
THE RESULTS INDICATE THAT	T	VP	0.0	15.1	**4.8**	**<.05**	0.0	2.2	1.2	not	6.6	22.3	**4.3**	**<.05**	7.4	31.6	**6.9**	**<.01**
THE RESULTS INDICATE THAT THE	T	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	2.2	5.6	0.7	not	4.9	6.3	0.1	not
THE RESULTS INDICATED THAT	T	VP	9.3	0.0	3.2	not	0.0	2.2	1.2	not	2.2	14.9	**5.1**	**<.05**	4.9	8.4	0.4	not
THE RESULTS OF A	T	NP	23.1	22.6	0.0	not	5.3	11.1	0.8	not	2.2	27.9	**12.4**	**<.001**	12.3	8.4	0.3	not
THE RESULTS OF OUR	T	NP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	4.4	18.6	**4.5**	**<.05**	12.3	14.8	0.1	not
THE RESULTS OF THE	T	NP	41.6	33.9	0.2	not	21.3	78.0	**13.8**	**<.001**	66.2	208.1	**37.4**	**<0.0001**	76.6	94.8	0.9	not
THE RESULTS OF THE CURRENT STUDY	T	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.9	14.8	2.2	not
THE RESULTS OF THE PRESENT STUDY	T	NP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	24.7	10.5	2.6	not
THE RESULTS OF THESE	T	NP	4.6	11.3	0.7	not	0.0	15.6	**8.5**	**<.01**	2.2	13.0	**4.1**	**<.05**	7.4	0.0	**4.7**	**<.05**
THE RESULTS OF THIS	T	NP	13.9	18.8	0.2	not	0.0	11.1	**6.1**	**<.05**	8.8	42.7	**11.7**	**<.001**	37.0	118.0	**19.2**	**<0.0001**
THE RESULTS OF THIS ANALYSIS ARE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	6.6	9.3	0.2	not	0.0	0.0	0.0	not
THE RESULTS OF THIS ANALYSIS ARE PRESENTED IN TABLE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	7.4	1.4	not	0.0	0.0	0.0	not
THE RESULTS OF THIS STUDY	T	NP	9.3	7.5	0.0	not	0.0	0.0	0.0	not	0.0	5.6	3.7	not	29.6	99.1	**17.0**	**<0.0001**
THE RESULTS OF THIS STUDY SUGGEST THAT	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	14.8	**4.2**	**<.05**
THE RESULTS SHOW THAT	T	VP	9.3	7.5	0.0	not	2.7	4.5	0.2	not	22.1	39.0	2.3	not	9.9	29.5	**4.4**	**<.05**
THE RESULTS SHOW THAT THE	T	VP	9.3	3.8	0.6	not	2.7	2.2	0.0	not	15.4	7.4	1.4	not	4.9	16.9	3.0	not
THE RESULTS SHOWED THAT	T	VP	18.5	3.8	2.6	not	2.7	2.2	0.0	not	15.4	13.0	0.1	not	17.3	6.3	2.3	not
THE RESULTS SHOWED THAT THE	T	VP	9.3	0.0	3.2	not	0.0	2.2	1.2	not	6.6	5.6	0.0	not	9.9	0.0	**6.2**	**<.05**
THE RESULTS SUGGEST THAT	T	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	2.2	9.3	2.3	not	2.5	23.2	**8.2**	**<.01**

Phrase																	
THE RIGHT SIDE OF THE	R	NP	4.6	0.0	1.6	not	2.7	4.5	0.2	not	4.4	3.7	0.0	not	2.5	0.0	1.6	not
THE RIGHT HAND SIDE OF	R	NP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	15.4	1.9	**6.2**	**<.05**	0.0	0.0	0.0	not
THE RISE OF THE	R	NP	0.0	18.8	**6.0**	**<.05**	0.0	0.0	0.0	not	0.0	7.4	**4.9**	**<.05**	0.0	10.5	**6.2**	**<.05**
THE ROLE OF THE	R	NP	23.1	18.8	0.1	not	2.7	6.7	0.7	not	15.4	9.3	0.8	not	24.7	42.2	2.0	not
THE ROTATION OF THE	R	NP	23.1	0.0	**8.0**	**<.01**	5.3	0.0	3.1	not	24.3	1.9	**11.6**	**<.001**	19.8	0.0	**12.4**	**<.001**
THE SAME AS THE	T	NP	4.6	0.0	1.6	not	5.3	4.5	0.0	not	17.6	3.7	**5.0**	**<.05**	12.3	0.0	**7.8**	**<.01**
THE SAME SET OF	R	NP	9.3	0.0	3.2	not	5.3	15.6	2.1	not	13.2	9.3	0.3	not	0.0	8.4	**4.9**	**<.05**
THE SAME WAY AS	T	NP	4.6	7.5	0.2	not	2.7	6.7	0.7	not	0.0	9.3	**6.1**	**<.05**	7.4	6.3	0.0	not
THE SCOPE OF THE	T	NP	0.0	3.8	1.2	not	2.7	4.5	0.2	not	2.2	3.7	0.2	not	12.3	19.0	0.6	not
THE SECOND PART OF THE	T	NP	0.0	3.8	1.2	not	0.0	15.6	**8.5**	**<.01**	0.0	0.0	0.0	not	2.5	2.1	0.0	not
THE SECOND RESEARCH QUESTION	T	NP	0.0	0.0	0.0	not	0.0	11.1	**6.1**	**<.05**	0.0	16.7	**11.0**	**<.001**	0.0	14.8	**8.6**	**<.01**
THE SHAPE OF THE	R	NP	0.0	3.8	1.2	not	8.0	0.0	**4.7**	**<.05**	26.5	5.6	**7.4**	**<.01**	12.3	2.1	3.6	not
THE SIGNIFICANCE OF THE	R	NP	4.6	3.8	2.0	not	24.0	8.9	3.0	not	2.2	35.3	**16.8**	**<0.0001**	4.9	16.9	3.0	not
THE SIZE OF THE	R	NP	32.4	26.4	0.1	not	34.6	53.5	1.7	not	35.3	29.7	0.2	not	49.4	33.7	1.3	not
THE SLOPE OF THE	R	NP	4.6	0.0	1.6	not	13.3	6.7	0.9	not	24.3	9.3	3.5	not	12.3	2.1	3.6	not
THE SQUARE ROOT OF	R	NP	4.6	0.0	1.6	not	0.0	13.4	**7.3**	**<.01**	6.6	13.0	1.0	not	9.9	0.0	**6.2**	**<.05**
THE SQUARE ROOT OF THE	R	NP	4.6	0.0	1.6	not	0.0	8.9	**4.9**	**<.05**	6.6	9.3	0.2	not	2.5	0.0	1.6	not
THE STANDARD DEVIATION OF	R	NP	0.0	0.0	0.0	not	10.6	15.6	0.4	not	13.2	13.0	0.0	not	0.0	0.0	0.0	not
THE STANDARD DEVIATION OF THE	R	NP	0.0	0.0	0.0	not	8.0	2.2	1.4	not	11.0	9.3	0.1	not	0.0	0.0	0.0	not
THE START OF THE	R	NP	0.0	3.8	1.2	not	21.3	20.0	0.0	not	19.9	11.1	1.2	not	4.9	10.5	0.9	not
THE START OF THE STUDY	R	NP	0.0	0.0	0.0	not	5.3	6.7	0.1	not	11.0	0.0	**7.8**	**<.01**	4.9	0.0	3.1	not
THE STRENGTH OF THE	R	NP	46.3	11.3	**5.6**	**<.05**	26.6	4.5	**7.3**	**<.01**	33.1	31.6	0.0	not	51.9	31.6	2.2	not
THE STRENGTH OF THE RELATIONSHIP BETWEEN	R	NP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.2	0.0	1.6	not	2.5	4.2	0.2	not
THE STRUCTURE OF THE	R	NP	69.4	7.5	**14.1**	**<.001**	21.3	24.5	0.1	not	19.9	3.7	**6.1**	**<.05**	24.7	8.4	3.7	not
THE SUM OF THE	R	NP	0.0	0.0	0.0	not	18.6	29.0	0.9	not	6.6	5.6	0.0	not	14.8	0.0	**9.3**	**<.01**
THE SURFACE OF THE	R	NP	46.3	0.0	**16.0**	**<0.0001**	21.3	0.0	**12.6**	**<.001**	26.5	0.0	**18.8**	**<0.0001**	37.0	0.0	**23.3**	**<0.0001**
THE THICKNESS OF THE	R	NP	13.9	0.0	**4.8**	**<.05**	24.0	0.0	**14.1**	**<.001**	35.3	0.0	**25.0**	**<0.0001**	29.6	0.0	**18.6**	**<0.0001**
THE THREE TYPES OF	R	NP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	8.8	20.4	2.3	not	7.4	10.5	0.2	not
THE TIME OF THE	R	NP	0.0	7.5	2.4	not	13.3	60.1	**13.0**	**<.001**	6.6	11.1	0.6	not	2.5	6.3	0.8	not
THE TIME OF THE STUDY	R	NP	0.0	3.8	1.2	not	2.7	11.1	2.2	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
THE TOP OF THE	R	NP	9.3	3.8	0.6	not	31.9	4.5	**9.8**	**<.01**	8.8	3.7	1.1	not	9.9	2.1	2.4	not
THE TOTAL AMOUNT OF	R	NP	9.3	0.0	3.2	not	13.3	6.7	0.9	not	4.4	3.7	0.0	not	2.5	2.1	0.0	not
THE TOTAL NUMBER OF	R	NP	0.0	7.5	2.4	not	53.2	106.9	**7.4**	**<.01**	24.3	46.5	3.5	not	9.9	4.2	1.0	not
THE TWO SETS OF	R	NP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	8.8	11.1	0.1	not	4.9	8.4	0.4	not
THE UNDERSTANDING OF THE	T	NP	18.5	3.8	2.6	not	2.7	6.7	0.7	not	2.2	0.0	1.6	not	7.4	16.9	1.6	not
THE USE OF A	R	NP	18.5	30.1	0.7	not	31.9	31.2	0.0	not	17.6	26.0	0.8	not	37.0	69.6	**4.4**	**<.05**
THE USE OF AN	R	NP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	0.0	9.3	**6.1**	**<.05**	9.9	31.6	**5.2**	**<.05**
THE USE OF THE	R	NP	27.8	18.8	0.4	not	5.3	26.7	**6.3**	**<.05**	6.6	27.9	**6.8**	**<.01**	27.2	29.5	0.0	not
THE VALIDITY OF THE	R	NP	32.4	15.1	1.6	not	0.0	24.5	**13.4**	**<.001**	2.2	18.6	**7.1**	**<.01**	12.3	10.5	0.1	not
THE VALUE OF A	R	NP	0.0	26.4	**8.3**	**<.01**	8.0	17.8	1.6	not	8.8	13.0	0.4	not	7.4	21.1	2.9	not
THE VALUE OF THE	R	NP	4.6	22.6	3.0	not	13.3	35.6	**4.3**	**<.05**	15.4	20.4	0.3	not	12.3	10.5	0.1	not
THE VALUES OF THE	R	NP	4.6	0.0	1.6	not	16.0	2.2	**4.9**	**<.05**	15.4	7.4	1.4	not	0.0	4.2	2.5	not
THE VARIABILITY IN THE	R	NP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	22.1	1.9	**10.2**	**<.01**	9.9	2.1	2.4	not
THE VAST MAJORITY OF	R	NP	4.6	18.8	2.2	not	5.3	2.2	0.5	not	8.8	13.0	0.4	not	4.9	8.4	0.4	not
THE WAY IN WHICH	R	NP	9.3	7.5	0.0	not	0.0	20.0	**10.9**	**<.001**	2.2	14.9	**5.1**	**<.05**	4.9	12.6	1.5	not
THE WAYS IN WHICH	R	NP	0.0	60.2	**19.1**	**<0.0001**	0.0	11.1	**6.1**	**<.05**	0.0	14.9	**9.8**	**<.01**	0.0	52.7	**30.9**	**<0.0001**
THERE APPEARS TO BE A	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	6.6	5.6	0.0	not	9.9	8.4	0.1	not
THERE ARE A NUMBER	R	VP	13.9	0.0	**4.8**	**<.05**	2.7	4.5	0.2	not	2.2	1.9	0.0	not	7.4	14.8	1.1	not
THERE ARE A NUMBER OF	R	NP_VP	13.9	0.0	**4.8**	**<.05**	2.7	4.5	0.2	not	2.2	1.9	0.0	not	7.4	14.8	1.1	not

Phrase																		
THERE IS A HIGH LEVEL OF	R	NP_VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	0.0	1.9	1.2	not	0.0	4.2	2.5	not
THERE IS A LACK OF	T	NP_VP	13.9	22.6	0.5	not	2.7	0.0	1.6	not	0.0	1.9	1.2	not	4.9	2.1	0.5	not
THERE IS A NEED FOR	P	NP_VP	4.6	26.4	**3.9**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
THERE IS A NEED TO	P	NP_VP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	4.9	19.0	3.8	not
THERE IS A SIGNIFICANT	R	NP_VP	4.6	3.8	0.0	not	2.7	4.5	0.2	not	6.6	18.6	2.9	not	7.4	0.0	**4.7**	**<.05**
THERE IS ALSO A	T	VP	4.6	3.8	0.0	not	2.7	0.0	1.6	not	13.2	18.6	0.4	not	17.3	8.4	1.4	not
THERE IS NO EVIDENCE	P	NP_VP	4.6	3.8	0.0	not	0.0	0.0	0.0	not	6.6	16.7	2.2	not	4.9	4.2	0.0	not
THERE IS NO EVIDENCE THAT	P	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	13.0	2.2	not	2.5	0.0	1.6	not
THERE IS NO REASON TO	P	VP	0.0	3.8	1.2	not	0.0	4.5	2.4	not	0.0	0.0	0.0	not	2.5	8.4	1.5	not
THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN	R	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	4.4	7.4	0.4	not	0.0	0.0	0.0	not
THERE WAS A SIGNIFICANT	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	37.5	14.9	**5.0**	**<.05**	0.0	0.0	0.0	not
THERE WAS NO SIGNIFICANT	R	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	50.7	18.6	**7.7**	**<.01**	17.3	0.0	**10.9**	**<.001**
THERE WAS NO SIGNIFICANT DIFFERENCE BETWEEN	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	5.6	1.6	not	2.5	0.0	1.6	not
THERE WAS NO SIGNIFICANT DIFFERENCE IN	R	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	17.6	1.9	**7.5**	**<.01**	4.9	0.0	3.1	not
THERE WAS NO STATISTICAL DIFFERENCE	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	1.9	**4.9**	**<.05**	2.5	0.0	1.6	not
THERE WERE NO SIGNIFICANT	R	NP_VP	0.0	0.0	0.0	not	2.7	2.2	0.0	not	57.4	22.3	**8.0**	**<.01**	4.9	2.1	0.5	not
THERE WERE NO SIGNIFICANT DIFFERENCES BETWEEN	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	26.5	5.6	**7.4**	**<.01**	4.9	0.0	3.1	not
THERE WERE NO SIGNIFICANT DIFFERENCES BETWEEN THE	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	8.8	5.6	0.4	not	2.5	0.0	1.6	not
THERE WERE NO SIGNIFICANT DIFFERENCES IN	R	NP_VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	6.6	3.7	0.4	not	0.0	0.0	0.0	not
THERE WERE NO STATISTICALLY SIGNIFICANT DIFFERENCES	R	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	11.0	1.9	3.6	not	9.9	0.0	**6.2**	**<.05**
THERE WERE SIGNIFICANT DIFFERENCES BETWEEN	R	NP_VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	6.6	11.1	0.6	not	0.0	0.0	0.0	not
THESE FINDINGS ARE CONSISTENT WITH	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	5.6	3.7	not	7.4	6.3	0.0	not
THESE FINDINGS SUGGEST THAT	T	NP_VP	13.9	0.0	**4.8**	**<.05**	0.0	0.0	0.0	not	2.2	13.0	**4.1**	**<.05**	14.8	29.5	2.1	not
THESE RESULTS ARE CONSISTENT	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	17.6	14.9	0.1	not	14.8	10.5	0.3	not
THESE RESULTS ARE CONSISTENT WITH	T	NP_VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	17.6	14.9	0.1	not	14.8	10.5	0.3	not
THESE RESULTS ARE CONSISTENT WITH THE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	13.2	5.6	1.6	not	4.9	6.3	0.1	not
THESE RESULTS INDICATE THAT	T	NP_VP	4.6	3.8	0.0	not	0.0	0.0	0.0	not	55.2	11.1	**16.0**	**<0.0001**	24.7	12.6	1.7	not
THESE RESULTS INDICATE THAT THE	T	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	13.2	1.9	**4.9**	**<.05**	9.9	0.0	**6.2**	**<.05**
THESE RESULTS SUGGEST THAT	T	NP_VP	13.9	3.8	1.5	not	0.0	4.5	2.4	not	64.0	31.6	**5.6**	**<.05**	42.0	19.0	**3.9**	**<.05**
THESE RESULTS SUGGEST THAT THE	T	NP_VP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	17.6	7.4	2.1	not	19.8	4.2	**4.9**	**<.05**
THEY ARE LESS LIKELY TO	P	VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
THEY ARE LIKELY TO	P	VP	0.0	7.5	2.4	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	4.9	8.4	0.4	not
THEY ARE MORE LIKELY TO	P	VP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	0.0	9.3	**6.1**	**<.05**	0.0	12.6	**7.4**	**<.01**
THIS ARTICLE CONTRIBUTES TO THE LITERATURE	T	NP_VP	0.0	15.1	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	4.2	2.5	not
THIS FINDING IS CONSISTENT WITH	T	NP_VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	9.3	**6.1**	**<.05**	4.9	19.0	3.8	not
THIS FINDING IS IN LINE WITH	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	5.6	3.7	not	0.0	8.4	**4.9**	**<.05**
THIS IS BECAUSE THE	T	VP	9.3	7.5	0.0	not	0.0	2.2	1.2	not	11.0	1.9	3.6	not	14.8	0.0	**9.3**	**<.01**
THIS IS CONSISTENT WITH	T	VP	0.0	3.8	1.2	not	2.7	4.5	0.2	not	39.7	29.7	0.7	not	49.4	27.4	2.8	not
THIS IS CONSISTENT WITH THE	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	37.5	13.0	**6.2**	**<.05**	29.6	6.3	**7.3**	**<.01**
THIS IS DUE TO THE	T	VP	4.6	3.8	0.0	not	0.0	0.0	0.0	not	11.0	3.7	1.9	not	17.3	2.1	**6.1**	**<.05**
THIS IS IN LINE	T	NP_VP	4.6	0.0	1.6	not	2.7	6.7	0.7	not	0.0	22.3	**14.7**	**<.001**	4.9	2.1	0.5	not
THIS IS IN LINE WITH	T	NP_VP	4.6	0.0	1.6	not	2.7	6.7	0.7	not	0.0	22.3	**14.7**	**<.001**	4.9	2.1	0.5	not
THIS IS IN LINE WITH THE	T	NP_VP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	0.0	7.4	**4.9**	**<.05**	2.5	2.1	0.0	not
THIS IS NOT THE CASE	T	NP_VP	4.6	3.8	0.0	not	2.7	2.2	0.0	not	6.6	16.7	2.2	not	14.8	10.5	0.3	not
THIS IS NOT THE CASE FOR	T	NP_VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	6.6	7.4	0.0	not	7.4	0.0	**4.7**	**<.05**
THIS IS NOT TO SAY THAT	T	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	8.4	**4.9**	**<.05**
THIS IS THE CASE	T	NP_VP	0.0	3.8	1.2	not	2.7	6.7	0.7	not	2.2	16.7	**6.1**	**<.05**	7.4	21.1	2.9	not
THIS IS THE FIRST	P	NP_VP	13.9	7.5	0.5	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	34.6	12.6	**4.7**	**<.05**
THIS IS THE FIRST STUDY TO	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	9.9	4.2	1.0	not

THIS LEADS TO THE	T	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	4.9	0.0	3.1	not
THIS MAY BE DUE TO THE	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	2.2	5.6	0.7	not	4.9	2.1	0.5	not
THIS MEANS THAT THE	T	VP	0.0	15.1	**4.8**	**<.05**	0.0	6.7	3.6	not	30.9	14.9	2.8	not	7.4	4.2	0.4	not
THIS PAPER IS TO	T	VP	13.9	7.5	0.5	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	2.1	1.2	not
THIS RESEARCH CONTRIBUTES TO THE	T	VP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	16.9	**9.9**	**<.01**
THIS RESULT IS CONSISTENT	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	11.0	7.4	0.3	not	9.9	12.6	0.1	not
THIS RESULT IS CONSISTENT WITH	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	11.0	7.4	0.3	not	9.9	12.6	0.1	not
THIS RESULT IS CONSISTENT WITH THE	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	8.8	3.7	1.1	not	4.9	2.1	0.5	not
THIS RESULT SUGGESTS THAT	T	VP	0.0	0.0	0.0	not	0.0	2.2	1.2	not	8.8	9.3	0.0	not	24.7	4.2	**7.2**	**<.01**
THIS STUDY CONTRIBUTES TO	T	VP	0.0	33.9	**10.7**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	21.1	**12.3**	**<.001**
THIS STUDY CONTRIBUTES TO THE	T	VP	0.0	22.6	**7.1**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	14.8	**8.6**	**<.01**
THIS STUDY IS THAT	T	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	12.3	37.9	**5.9**	**<.05**
THIS STUDY IS THE FIRST TO	P	VP	0.0	11.3	3.6	not	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	6.3	3.7	not
THIS STUDY IS TO	T	VP	50.9	56.5	0.1	not	5.3	8.9	0.4	not	0.0	0.0	0.0	not	0.0	6.3	3.7	not
THIS STUDY SHOWS THAT	T	VP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	7.4	29.5	**6.1**	**<.05**
THIS STUDY WAS TO	T	VP	87.9	7.5	**19.6**	**<0.0001**	0.0	4.5	2.4	not	2.2	5.6	0.7	not	7.4	16.9	1.6	not
THIS SUGGESTS THAT THE	T	VP	0.0	11.3	3.6	not	5.3	2.2	0.5	not	19.9	26.0	0.4	not	32.1	23.2	0.6	not
THIS WAS NOT THE CASE	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	9.3	**6.1**	**<.05**	2.5	14.8	**4.2**	**<.05**
THROUGH THE USE OF	R	PP	27.8	33.9	0.1	not	0.0	17.8	**9.7**	**<.01**	0.0	9.3	**6.1**	**<.05**	4.9	19.0	3.8	not
TO A BETTER UNDERSTANDING OF	T	PP	0.0	15.1	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	6.3	0.8	not
TO A BETTER UNDERSTANDING OF THE	T	PP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	2.5	4.2	0.2	not
TO A GREATER EXTENT THAN	R	PP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	2.2	9.3	2.3	not	4.9	0.0	3.1	not
TO A LESSER EXTENT	R	PP	9.3	0.0	3.2	not	0.0	4.5	2.4	not	26.5	31.6	0.2	not	17.3	10.5	0.7	not
TO A WIDE RANGE OF	R	PP	13.9	3.8	1.5	not	2.7	0.0	1.6	not	2.2	0.0	1.6	not	0.0	4.2	2.5	not
TO ACCOUNT FOR THE	R	VP	13.9	7.5	0.5	not	24.0	60.1	**6.5**	**<.05**	6.6	14.9	1.6	not	17.3	23.2	0.4	not
TO AN INCREASE IN	R	PP	13.9	3.8	1.5	not	0.0	0.0	0.0	not	8.8	1.9	2.5	not	22.2	8.4	2.8	not
TO ASSESS THE EXTENT TO WHICH	R	NP_VP	0.0	3.8	1.2	not	0.0	11.1	**6.1**	**<.05**	0.0	1.9	1.2	not	0.0	0.0	0.0	not
TO ASSESS THE IMPACT OF	R	NP_VP	4.6	7.5	0.2	not	0.0	8.9	**4.9**	**<.05**	0.0	0.0	0.0	not	2.5	0.0	1.6	not
TO BE ABLE TO	P	NP_VP	13.9	3.8	1.5	not	5.3	15.6	2.1	not	15.4	7.4	1.4	not	9.9	23.2	2.4	not
TO BE ASSOCIATED WITH	T	VP	13.9	11.3	0.1	not	2.7	6.7	0.7	not	11.0	20.4	1.4	not	27.2	4.2	**8.4**	**<.01**
TO BE DUE TO	P	VP	9.3	0.0	3.2	not	0.0	2.2	1.2	not	28.7	1.9	**14.4**	**<.001**	27.2	0.0	**17.1**	**<0.0001**
TO BE INCLUDED IN THE	R	NP_VP	0.0	3.8	1.2	not	8.0	13.4	0.6	not	0.0	5.6	3.7	not	0.0	4.2	2.5	not
TO BE MORE LIKELY TO	P	NP_VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	5.6	3.7	not	0.0	4.2	2.5	not
TO BE THE MOST	P	NP_VP	23.1	7.5	2.0	not	0.0	6.7	3.6	not	8.8	14.9	0.8	not	9.9	6.3	0.3	not
TO BETTER UNDERSTAND THE	T	VP	9.3	33.9	3.5	not	5.3	8.9	0.4	not	6.6	3.7	0.4	not	17.3	25.3	0.7	not
TO CONTRIBUTE TO THE	T	VP	13.9	45.2	**4.1**	**<.05**	0.0	8.9	**4.9**	**<.05**	2.2	13.0	**4.1**	**<.05**	7.4	8.4	0.0	not
TO CONTROL FOR THE	R	VP	0.0	3.8	1.2	not	2.7	73.5	**32.7**	**<0.0001**	0.0	5.6	3.7	not	0.0	6.3	3.7	not
TO DETERMINE IF THE	R	VP	4.6	7.5	0.2	not	5.3	11.1	0.8	not	17.6	7.4	2.1	not	4.9	4.2	0.0	not
TO DETERMINE WHETHER THE	R	VP	9.3	3.8	0.6	not	8.0	20.0	2.2	not	17.6	16.7	0.0	not	4.9	10.5	0.9	not
TO DIFFERENCES IN THE	R	PP	4.6	7.5	0.2	not	0.0	0.0	0.0	not	4.4	5.6	0.1	not	17.3	2.1	**6.1**	**<.05**
TO ENSURE THAT THE	R	VP	18.5	3.8	2.6	not	13.3	44.6	**7.2**	**<.01**	6.6	3.7	0.4	not	4.9	10.5	0.9	not
TO EXAMINE THE EFFECT OF	R	NP_VP	9.3	3.8	0.6	not	10.6	6.7	0.4	not	0.0	5.6	3.7	not	0.0	6.3	3.7	not
TO EXAMINE THE EFFECTS OF	T	NP_VP	4.6	7.5	0.2	not	5.3	4.5	0.0	not	0.0	1.9	1.2	not	0.0	0.0	0.0	not
TO EXAMINE THE IMPACT OF	R	NP_VP	0.0	3.8	1.2	not	0.0	15.6	**8.5**	**<.01**	0.0	0.0	0.0	not	0.0	4.2	2.5	not
TO EXAMINE THE RELATIONSHIP BETWEEN	R	NP_VP	0.0	15.1	**4.8**	**<.05**	5.3	2.2	0.5	not	0.0	5.6	3.7	not	0.0	2.1	1.2	not
TO EXAMINE WHETHER THE	R	VP	0.0	7.5	2.4	not	5.3	13.4	1.4	not	0.0	18.6	**12.2**	**<.001**	2.5	4.2	0.2	not
TO FOCUS ON THE	T	VP	9.3	11.3	0.0	not	2.7	0.0	1.6	not	0.0	9.3	**6.1**	**<.05**	0.0	16.9	**9.9**	**<.01**
TO GAIN INSIGHT INTO THE	T	NP_VP	13.9	3.8	1.5	not	0.0	2.2	1.2	not	4.4	0.0	3.1	not	4.9	0.0	3.1	not
TO HIGHER LEVELS OF	R	PP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	2.2	0.0	1.6	not	4.9	12.6	1.5	not

TO INVESTIGATE THE EFFECT OF	R	NP_VP	18.5	0.0	**6.4**	**<.05**	8.0	0.0	**4.7**	**<.05**	11.0	0.0	**7.8**	**<.01**	7.4	2.1	1.4	not
TO INVESTIGATE THE EFFECTS OF	R	NP_VP	18.5	0.0	**6.4**	**<.05**	8.0	4.5	0.4	not	2.2	1.9	0.0	not	0.0	2.1	1.2	not
TO INVESTIGATE THE INFLUENCE OF	R	NP_VP	9.3	0.0	3.2	not	5.3	0.0	3.1	not	2.2	0.0	1.6	not	4.9	2.1	0.5	not
TO ONE OF THE	R	PP	0.0	3.8	1.2	not	8.0	22.3	2.8	not	4.4	13.0	2.2	not	2.5	6.3	0.8	not
TO OUR UNDERSTANDING OF	R	PP	9.3	22.6	1.4	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	25.3	**14.8**	**<.001**
TO PARTICIPATE IN THE	R	VP	4.6	11.3	0.7	not	8.0	46.8	**12.2**	**<.001**	0.0	11.1	**7.3**	**<.01**	7.4	12.6	0.6	not
TO PARTICIPATE IN THE STUDY	R	NP_VP	0.0	0.0	0.0	not	8.0	29.0	**5.1**	**<.05**	0.0	0.0	0.0	not	0.0	6.3	3.7	not
TO PLAY AN IMPORTANT ROLE IN	P	NP_VP	4.6	3.8	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.9	4.2	0.0	not
TO PROVIDE A MORE	R	VP	0.0	11.3	3.6	not	5.3	8.9	0.4	not	2.2	5.6	0.7	not	4.9	8.4	0.4	not
TO SHED LIGHT ON THE	T	VP	4.6	18.8	2.2	not	0.0	2.2	1.2	not	2.2	1.9	0.0	not	0.0	4.2	2.5	not
TO TAKE ADVANTAGE OF	T	NP_VP	9.3	3.8	0.6	not	2.7	4.5	0.2	not	0.0	3.7	2.4	not	2.5	6.3	0.8	not
TO TAKE INTO ACCOUNT	R	VP	0.0	7.5	2.4	not	16.0	26.7	1.1	not	4.4	5.6	0.1	not	2.5	12.6	3.2	not
TO TAKE INTO ACCOUNT THE	R	VP	0.0	3.8	1.2	not	10.6	6.7	0.4	not	2.2	0.0	1.6	not	2.5	10.5	2.3	not
TO TEST OUR HYPOTHESES	R	NP_VP	0.0	3.8	1.2	not	0.0	31.2	**17.0**	**<0.0001**	0.0	9.3	**6.1**	**<.05**	0.0	0.0	0.0	not
TO TEST THE HYPOTHESES	T	NP_VP	0.0	3.8	1.2	not	0.0	31.2	**17.0**	**<0.0001**	0.0	14.9	**9.8**	**<.01**	0.0	2.1	1.2	not
TO THAT OF THE	R	PP	32.4	0.0	**11.2**	**<.001**	8.0	2.2	1.4	not	59.6	1.9	**34.9**	**<0.0001**	44.5	4.2	**17.4**	**<0.0001**
TO THE AMOUNT OF	R	PP	4.6	11.3	0.7	not	5.3	4.5	0.0	not	6.6	3.7	0.4	not	7.4	10.5	0.2	not
TO THE BEST OF	P	PP	9.3	18.8	0.8	not	0.0	2.2	1.2	not	0.0	0.0	0.0	not	22.2	14.8	0.7	not
TO THE BEST OF OUR KNOWLEDGE	P	PP	4.6	15.1	1.4	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	14.8	8.4	0.8	not
TO THE DEVELOPMENT OF	R	PP	27.8	11.3	1.7	not	0.0	4.5	2.4	not	2.2	7.4	1.4	not	14.8	27.4	1.7	not
TO THE EFFECT OF	R	PP	9.3	7.5	0.0	not	0.0	2.2	1.2	not	13.2	9.3	0.3	not	4.9	2.1	0.5	not
TO THE EXTENT THAT	T	PP	0.0	33.9	**10.7**	**<.01**	0.0	13.4	**7.3**	**<.01**	2.2	11.1	3.2	not	4.9	63.2	**25.2**	**<0.0001**
TO THE EXTENT THAT THE	T	PP	0.0	7.5	2.4	not	0.0	6.7	3.6	not	2.2	3.7	0.2	not	2.5	16.9	**5.1**	**<.05**
TO THE FACT THAT	P	PP	4.6	11.3	0.7	not	2.7	13.4	3.1	not	24.3	31.6	0.5	not	37.0	31.6	0.2	not
TO THE FORMATION OF	R	PP	41.6	3.8	**9.1**	**<.01**	5.3	2.2	0.5	not	30.9	0.0	**21.9**	**<0.0001**	51.9	0.0	**32.6**	**<0.0001**
TO THE FORMATION OF A	R	PP	9.3	0.0	3.2	not	2.7	0.0	1.6	not	11.0	0.0	**7.8**	**<.01**	17.3	0.0	**10.9**	**<.001**
TO THE IMPORTANCE OF	P	PP	4.6	7.5	0.2	not	2.7	6.7	0.7	not	2.2	7.4	1.4	not	0.0	21.1	**12.3**	**<.001**
TO THE IMPORTANCE OF THE	P	PP	0.0	0.0	0.0	not	2.7	4.5	0.2	not	0.0	3.7	2.4	not	0.0	2.1	1.2	not
TO THE LEVEL OF	T	PP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	0.0	11.1	**7.3**	**<.01**	0.0	12.6	**7.4**	**<.01**
TO THE LITERATURE ON	T	PP	0.0	33.9	**10.7**	**<.01**	0.0	0.0	0.0	not	0.0	1.9	1.2	not	0.0	25.3	**14.8**	**<.001**
TO THE NUMBER OF	R	PP	4.6	18.8	2.2	not	8.0	29.0	**5.1**	**<.05**	8.8	9.3	0.0	not	12.3	14.8	0.1	not
TO THE RELATIONSHIP BETWEEN	T	PP	0.0	18.8	**6.0**	**<.05**	0.0	0.0	0.0	not	0.0	7.4	**4.9**	**<.05**	0.0	16.9	**9.9**	**<.01**
TO THE REST OF THE	R	PP	0.0	3.8	1.2	not	5.3	6.7	0.1	not	4.4	1.9	0.5	not	2.5	2.1	0.0	not
TO THE STUDY OF	T	PP	4.6	37.6	**6.8**	**<.01**	0.0	2.2	1.2	not	0.0	0.0	0.0	not	0.0	19.0	**11.1**	**<.001**
TO THE TOTAL NUMBER OF	R	PP	0.0	3.8	1.2	not	2.7	4.5	0.2	not	8.8	11.1	0.1	not	0.0	0.0	0.0	not
TO THE USE OF	R	PP	18.5	15.1	0.1	not	0.0	4.5	2.4	not	2.2	14.9	**5.1**	**<.05**	14.8	16.9	0.1	not
UNDERSTANDING OF THE RELATIONSHIP BETWEEN	T	NP	0.0	7.5	2.4	not	0.0	0.0	0.0	not	2.2	1.9	0.0	not	2.5	10.5	2.3	not
USED AS A MEASURE OF	R	NP_VP	0.0	0.0	0.0	not	10.6	4.5	1.1	not	0.0	3.7	2.4	not	0.0	0.0	0.0	not
USED IN THE ANALYSIS	R	NP_VP	0.0	3.8	1.2	not	5.3	24.5	**5.4**	**<.05**	2.2	3.7	0.2	not	4.9	2.1	0.5	not
USED IN THE CURRENT STUDY	R	NP_VP	0.0	0.0	0.0	not	5.3	13.4	1.4	not	2.2	0.0	1.6	not	7.4	0.0	**4.7**	**<.05**
USED IN THE PRESENT	R	NP_VP	9.3	0.0	3.2	not	13.3	6.7	0.9	not	4.4	0.0	3.1	not	7.4	12.6	0.6	not
USED IN THE PRESENT STUDY	R	NP_VP	0.0	0.0	0.0	not	8.0	6.7	0.0	not	2.2	0.0	1.6	not	4.9	10.5	0.9	not
USED IN THIS STUDY	R	NP_VP	27.8	3.8	**5.1**	**<.05**	130.4	53.5	**13.8**	**<.001**	19.9	13.0	0.7	not	27.2	21.1	0.3	not
USED IN THIS STUDY ARE	R	NP_VP	4.6	0.0	1.6	not	26.6	2.2	**10.2**	**<.01**	2.2	0.0	1.6	not	0.0	2.1	1.2	not
USED IN THIS STUDY IS	R	NP_VP	0.0	0.0	0.0	not	10.6	8.9	0.1	not	2.2	0.0	1.6	not	0.0	2.1	1.2	not
USED IN THIS STUDY WERE	R	NP_VP	0.0	0.0	0.0	not	21.3	4.5	**5.0**	**<.05**	2.2	0.0	1.6	not	4.9	0.0	3.1	not
USED TO CALCULATE THE	R	VP	0.0	0.0	0.0	not	42.6	4.5	**15.0**	**<.001**	2.2	5.6	0.7	not	12.3	2.1	3.6	not
USED TO DETERMINE THE	R	VP	9.3	0.0	3.2	not	39.9	4.5	**13.7**	**<.001**	8.8	0.0	**6.3**	**<.05**	12.3	0.0	**7.8**	**<.01**
USED TO ESTIMATE THE	R	VP	0.0	0.0	0.0	not	13.3	6.7	0.9	not	11.0	3.7	1.9	not	7.4	0.0	**4.7**	**<.05**

USED TO MEASURE THE	R	VP	13.9	3.8	1.5	not	29.3	13.4	2.5	not	6.6	5.6	0.0	not	0.0	2.1	1.2	not
USED TO TEST THE	R	VP	0.0	7.5	2.4	not	10.6	15.6	0.4	not	0.0	13.0	**8.6**	**<.01**	0.0	4.2	2.5	not
USING DATA FROM THE	R	VP	4.6	15.1	1.4	not	0.0	11.1	**6.1**	**<.05**	0.0	7.4	**4.9**	**<.05**	0.0	2.1	1.2	not
VARIABLES USED IN THE	R	NP_VP	0.0	3.8	1.2	not	5.3	20.0	3.7	not	0.0	13.0	**8.6**	**<.01**	0.0	6.3	3.7	not
WAS ADDED TO THE	R	VP	0.0	0.0	0.0	not	82.5	6.7	**32.1**	**<0.0001**	6.6	11.1	0.6	not	2.5	0.0	1.6	not
WAS BASED ON THE	T	VP	4.6	3.8	0.0	not	29.3	20.0	0.7	not	2.2	7.4	1.4	not	4.9	4.2	0.0	not
WAS CARRIED OUT IN	R	VP	0.0	3.8	1.2	not	34.6	11.1	**5.2**	**<.05**	0.0	0.0	0.0	not	0.0	2.1	1.2	not
WAS CARRIED OUT USING A	R	VP	0.0	0.0	0.0	not	24.0	2.2	**8.9**	**<.01**	0.0	1.9	1.2	not	0.0	0.0	0.0	not
WAS CONSIDERED TO BE	R	VP	0.0	3.8	1.2	not	24.0	8.9	3.0	not	4.4	3.7	0.0	not	2.5	0.0	1.6	not
WAS FOUND THAT THE	P	VP	9.3	0.0	3.2	not	8.0	2.2	1.4	not	15.4	3.7	**3.9**	**<.05**	7.4	4.2	0.4	not
WAS FOUND TO BE	P	VP	27.8	3.8	**5.1**	**<.05**	8.0	11.1	0.2	not	24.3	20.4	0.2	not	37.0	16.9	3.4	not
WAS MEASURED BY THE	R	VP	0.0	3.8	1.2	not	18.6	42.3	3.8	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
WAS MEASURED USING A	R	VP	0.0	0.0	0.0	not	50.6	26.7	3.1	not	0.0	0.0	0.0	not	2.5	0.0	1.6	not
WAS OBSERVED IN THE	R	VP	4.6	0.0	1.6	not	2.7	2.2	0.0	not	39.7	3.7	**17.6**	**<0.0001**	12.3	2.1	3.6	not
WAS POSITIVELY RELATED TO	R	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	37.2	**24.4**	**<0.0001**	4.9	10.5	0.9	not
WAS SIMILAR TO THAT	T	VP	0.0	0.0	0.0	not	5.3	4.5	0.0	not	24.3	1.9	**11.6**	**<.001**	9.9	0.0	**6.2**	**<.05**
WAS USED AS A	R	VP	4.6	7.5	0.2	not	74.5	22.3	**12.4**	**<.001**	8.8	3.7	1.1	not	4.9	2.1	0.5	not
WAS USED AS THE	R	VP	0.0	0.0	0.0	not	58.6	24.5	**6.0**	**<.05**	13.2	3.7	2.8	not	0.0	0.0	0.0	not
WAS USED FOR THE	R	VP	0.0	0.0	0.0	not	45.2	6.7	**13.5**	**<.001**	13.2	1.9	**4.9**	**<.05**	17.3	0.0	**10.9**	**<.001**
WAS USED TO DETERMINE	R	VP	0.0	0.0	0.0	not	45.2	2.2	**20.2**	**<0.0001**	4.4	0.0	3.1	not	0.0	0.0	0.0	not
WAS USED TO DETERMINE THE	R	VP	0.0	0.0	0.0	not	21.3	0.0	**12.6**	**<.001**	2.2	0.0	1.6	not	0.0	0.0	0.0	not
WAS USED TO MEASURE	R	VP	0.0	0.0	0.0	not	24.0	17.8	0.4	not	2.2	0.0	1.6	not	0.0	0.0	0.0	not
WAS USED TO MEASURE THE	R	VP	0.0	0.0	0.0	not	18.6	8.9	1.4	not	2.2	0.0	1.6	not	0.0	0.0	0.0	not
WE ALSO FOUND THAT	T	VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	4.4	7.4	0.4	not	17.3	12.6	0.3	not
WE ARE ABLE TO	P	VP	4.6	7.5	0.2	not	2.7	20.0	**6.0**	**<.05**	2.2	1.9	0.0	not	2.5	19.0	**6.2**	**<.05**
WE ARE INTERESTED IN	T	VP	4.6	11.3	0.7	not	2.7	13.4	3.1	not	2.2	5.6	0.7	not	2.5	2.1	0.0	not
WE ARGUE THAT THE	T	VP	0.0	33.9	**10.7**	**<.01**	0.0	6.7	3.6	not	0.0	3.7	2.4	not	0.0	19.0	**11.1**	**<.001**
WE BELIEVE THAT THE	P	VP	0.0	3.8	1.2	not	0.0	6.7	3.6	not	6.6	0.0	**4.7**	**<.05**	4.9	27.4	**7.4**	**<.01**
WE CAN SEE THAT	P	VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	4.4	18.6	**4.5**	**<.05**	4.9	10.5	0.9	not
WE CAN SEE THAT THE	P	VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	4.4	11.1	1.5	not	4.9	0.0	3.1	not
WE CANNOT EXCLUDE THE POSSIBILITY THAT	P	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	4.4	1.9	0.5	not	9.9	0.0	**6.2**	**<.05**
WE CONTROLLED FOR THE	R	VP	0.0	0.0	0.0	not	0.0	51.2	**28.0**	**<0.0001**	0.0	13.0	**8.6**	**<.01**	0.0	4.2	2.5	not
WE EXPECT THAT THE	T	VP	0.0	11.3	3.6	not	0.0	4.5	2.4	not	2.2	1.9	0.0	not	4.9	4.2	0.0	not
WE FIND THAT THE	T	VP	4.6	11.3	0.7	not	2.7	0.0	1.6	not	6.6	16.7	2.2	not	9.9	40.0	**8.4**	**<.01**
WE FOCUS ON THE	T	VP	13.9	52.7	**5.6**	**<.05**	5.3	8.9	0.4	not	6.6	3.7	0.4	not	4.9	4.2	0.0	not
WE FOUND THAT THE	T	VP	9.3	0.0	3.2	not	2.7	11.1	2.2	not	33.1	24.2	0.7	not	42.0	19.0	**3.9**	**<.05**
WE NOTE THAT THE	P	VP	4.6	0.0	1.6	not	0.0	4.5	2.4	not	15.4	5.6	2.4	not	19.8	6.3	3.2	not
WE PRESENT THE RESULTS OF	T	NP_VP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	0.0	9.3	**6.1**	**<.05**	0.0	0.0	0.0	not
WE PROPOSE THE FOLLOWING HYPOTHESIS	T	NP_VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	0.0	0.0	not
WE SEE THAT THE	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	17.6	31.6	2.0	not	2.5	2.1	0.0	not
WE WERE ABLE TO	P	VP	0.0	0.0	0.0	not	8.0	17.8	1.6	not	0.0	5.6	3.7	not	14.8	16.9	0.1	not
WE WERE NOT ABLE TO	P	VP	0.0	0.0	0.0	not	0.0	6.7	3.6	not	0.0	1.9	1.2	not	7.4	6.3	0.0	not
WERE EXCLUDED FROM THE	R	VP	0.0	0.0	0.0	not	39.9	17.8	3.6	not	2.2	5.6	0.7	not	0.0	0.0	0.0	not
WERE FOUND IN THE	P	VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	11.0	3.7	1.9	not	14.8	10.5	0.3	not
WERE FOUND TO BE	P	VP	27.8	3.8	**5.1**	**<.05**	16.0	4.5	2.9	not	30.9	9.3	**6.1**	**<.05**	22.2	12.6	1.2	not
WERE INCLUDED IN THE	R	VP	0.0	0.0	0.0	not	31.9	35.6	0.1	not	6.6	11.1	0.6	not	9.9	2.1	2.4	not
WERE INCLUDED IN THE ANALYSIS	R	NP_VP	0.0	0.0	0.0	not	10.6	8.9	0.1	not	0.0	0.0	0.0	not	2.5	0.0	1.6	not
WERE LESS LIKELY TO	P	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	8.8	22.3	2.9	not	2.5	8.4	1.5	not
WERE MORE LIKELY TO	P	VP	4.6	3.8	0.0	not	2.7	15.6	**4.1**	**<.05**	6.6	76.2	**32.9**	**<0.0001**	7.4	42.2	**11.5**	**<.001**

Phrase																		
WERE MORE LIKELY TO BE	P	VP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	2.2	11.1	3.2	not	4.9	2.1	0.5	not
WERE NO SIGNIFICANT DIFFERENCES BETWEEN	R	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	26.5	5.6	**7.4**	**<.01**	4.9	0.0	3.1	not
WERE NO SIGNIFICANT DIFFERENCES BETWEEN THE	R	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	8.8	5.6	0.4	not	2.5	0.0	1.6	not
WERE NOT INCLUDED IN THE	R	VP	0.0	0.0	0.0	not	5.3	6.7	0.1	not	0.0	3.7	2.4	not	0.0	4.2	2.5	not
WERE OBTAINED FROM THE	R	VP	0.0	0.0	0.0	not	66.5	11.1	**18.4**	**<0.0001**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
WERE RANDOMLY ASSIGNED TO ONE OF	R	NP_VP	0.0	0.0	0.0	not	0.0	17.8	**9.7**	**<.01**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
WERE REMOVED FROM THE	R	VP	0.0	0.0	0.0	not	31.9	15.6	2.4	not	2.2	7.4	1.4	not	0.0	0.0	0.0	not
WERE SIGNIFICANTLY MORE LIKELY TO	R	VP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	0.0	27.9	**18.3**	**<0.0001**	2.5	2.1	0.0	not
WERE SIMILAR TO THOSE	T	VP	4.6	3.8	0.0	not	0.0	4.5	2.4	not	13.2	7.4	0.8	not	17.3	0.0	**10.9**	**<.001**
WERE USED IN THIS STUDY	R	NP_VP	4.6	0.0	1.6	not	18.6	2.2	**6.2**	**<.05**	0.0	0.0	0.0	not	0.0	0.0	0.0	not
WHEN COMPARED TO THE	T	VP	13.9	0.0	**4.8**	**<.05**	8.0	2.2	1.4	not	26.5	7.4	**5.7**	**<.05**	14.8	2.1	**4.8**	**<.05**
WHEN COMPARED WITH THE	T	VP	4.6	0.0	1.6	not	0.0	2.2	1.2	not	24.3	5.6	**6.3**	**<.05**	17.3	4.2	3.8	not
WHEN IT COMES TO	P	VP	0.0	33.9	**10.7**	**<.01**	0.0	13.4	**7.3**	**<.01**	0.0	14.9	**9.8**	**<.01**	0.0	35.8	**21.0**	**<0.0001**
WHEN IT COMES TO THE	P	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	1.9	1.2	not	0.0	8.4	**4.9**	**<.05**
WHEN THERE IS A	T	VP	0.0	3.8	1.2	not	0.0	13.4	**7.3**	**<.01**	2.2	9.3	2.3	not	4.9	8.4	0.4	not
WHERE N IS THE NUMBER OF	R	NP_VP	0.0	0.0	0.0	not	16.0	6.7	1.6	not	0.0	0.0	0.0	not	2.5	0.0	1.6	not
WHETHER OR NOT THE	P	OTHER	4.6	15.1	1.4	not	10.6	40.1	**7.3**	**<.01**	6.6	13.0	1.0	not	0.0	25.3	**14.8**	**<.001**
WHETHER OR NOT TO	P	OTHER	0.0	7.5	2.4	not	2.7	6.7	0.7	not	0.0	9.3	**6.1**	**<.05**	2.5	4.2	0.2	not
WHICH IS CONSISTENT WITH	T	VP	0.0	0.0	0.0	not	5.3	2.2	0.5	not	41.9	16.7	**5.6**	**<.05**	39.5	12.6	**6.4**	**<.05**
WHICH IS CONSISTENT WITH THE	T	VP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	30.9	7.4	**7.7**	**<.01**	24.7	10.5	2.6	not
WHICH IS IN LINE WITH	T	VP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	5.6	3.7	not	2.5	4.2	0.2	not
WHICH IS RELATED TO THE	R	VP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	4.4	0.0	3.1	not	9.9	0.0	**6.2**	**<.05**
WHICH IS SIMILAR TO THE	T	VP	0.0	3.8	1.2	not	8.0	2.2	1.4	not	6.6	0.0	**4.7**	**<.05**	7.4	0.0	**4.7**	**<.05**
WHILE AT THE SAME TIME	T	PP	0.0	3.8	1.2	not	0.0	2.2	1.2	not	0.0	5.6	3.7	not	2.5	6.3	0.8	not
WILL BE DISCUSSED IN	T	VP	0.0	7.5	2.4	not	2.7	6.7	0.7	not	15.4	9.3	0.8	not	4.9	4.2	0.0	not
WILL BE LESS LIKELY TO	P	VP	0.0	0.0	0.0	not	0.0	4.5	2.4	not	0.0	5.6	3.7	not	0.0	2.1	1.2	not
WILL BE MORE LIKELY TO	P	VP	0.0	15.1	**4.8**	**<.05**	0.0	0.0	0.0	not	0.0	3.7	2.4	not	0.0	2.1	1.2	not
WITH A HIGH LEVEL OF	R	PP	0.0	3.8	1.2	not	2.7	4.5	0.2	not	0.0	3.7	2.4	not	0.0	4.2	2.5	not
WITH A WIDE RANGE OF	R	PP	4.6	3.8	0.0	not	2.7	2.2	0.0	not	0.0	5.6	3.7	not	0.0	4.2	2.5	not
WITH AN AVERAGE OF	R	PP	0.0	0.0	0.0	not	5.3	22.3	**4.5**	**<.05**	6.6	7.4	0.0	not	4.9	0.0	3.1	not
WITH AN INCREASE IN	R	PP	4.6	0.0	1.6	not	2.7	0.0	1.6	not	13.2	13.0	0.0	not	9.9	6.3	0.3	not
WITH AN INCREASE IN THE	R	PP	0.0	0.0	0.0	not	2.7	0.0	1.6	not	6.6	1.9	1.4	not	4.9	0.0	3.1	not
WITH HIGH LEVELS OF	R	PP	18.5	11.3	0.4	not	0.0	4.5	2.4	not	8.8	5.6	0.4	not	9.9	8.4	0.1	not
WITH HIGHER LEVELS OF	R	PP	0.0	7.5	2.4	not	2.7	6.7	0.7	not	2.2	27.9	**12.4**	**<.001**	0.0	23.2	**13.6**	**<.001**
WITH REGARD TO THE	T	PP	4.6	3.8	0.0	not	2.7	24.5	**8.1**	**<.01**	2.2	31.6	**14.6**	**<.001**	12.3	35.8	**5.2**	**<.05**
WITH RESPECT TO THE	T	PP	32.4	41.4	0.3	not	50.6	15.6	**8.1**	**<.01**	72.8	48.3	2.5	not	64.2	27.4	**6.7**	**<.01**
WITH THAT OF THE	R	PP	13.9	0.0	**4.8**	**<.05**	5.3	4.5	0.0	not	17.6	1.9	**7.5**	**<.01**	9.9	4.2	1.0	not
WITH THE EXCEPTION OF	T	PP	4.6	11.3	0.7	not	21.3	15.6	0.4	not	57.4	46.5	0.6	not	29.6	14.8	2.2	not
WITH THE EXCEPTION OF THE	T	PP	0.0	3.8	1.2	not	13.3	0.0	**7.9**	**<.01**	22.1	7.4	3.8	not	9.9	6.3	0.3	not
WITH THE FACT THAT	P	PP	0.0	7.5	2.4	not	0.0	2.2	1.2	not	17.6	9.3	1.3	not	12.3	16.9	0.3	not
WITH THE INCREASE OF	R	PP	0.0	0.0	0.0	not	0.0	0.0	0.0	not	55.2	0.0	**39.1**	**<0.0001**	61.7	2.1	**31.5**	**<0.0001**
WITH THE NUMBER OF	R	PP	0.0	7.5	2.4	not	8.0	4.5	0.4	not	13.2	22.3	1.1	not	0.0	0.0	0.0	not
WITH THE PRESENCE OF	R	PP	13.9	0.0	**4.8**	**<.05**	5.3	0.0	3.1	not	11.0	1.9	3.6	not	17.3	6.3	2.3	not
WITH THE RESULTS OF THE	T	PP	4.6	0.0	1.6	not	0.0	0.0	0.0	not	6.6	5.6	0.0	not	14.8	2.1	**4.8**	**<.05**
WITH THE USE OF	R	PP	13.9	15.1	0.0	not	5.3	0.0	3.1	not	11.0	11.1	0.0	not	7.4	10.5	0.2	not
WITHIN THE CONTEXT OF	T	PP	9.3	7.5	0.0	not	0.0	2.2	1.2	not	0.0	3.7	2.4	not	9.9	10.5	0.0	not
WOMEN ARE MORE LIKELY	P	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	24.2	**15.9**	**<0.0001**	0.0	6.3	3.7	not
WOMEN ARE MORE LIKELY TO	P	VP	0.0	3.8	1.2	not	0.0	0.0	0.0	not	0.0	18.6	**12.2**	**<.001**	0.0	6.3	3.7	not
WOULD BE EXPECTED TO	P	VP	9.3	0.0	3.2	not	0.0	4.5	2.4	not	4.4	0.0	3.1	not	19.8	6.3	3.2	not

| WOULD BE MORE LIKELY TO | P | VP | 0.0 | 0.0 | 0.0 | not | 0.0 | 0.0 | 0.0 | not | 0.0 | 9.3 | **6.1** | **<.05** | 0.0 | 6.3 | 3.7 | not |

Appendix C

Table C1. Full factorial structure matrix for the four-factor solution

	Factor			
Feature	**1**	**2**	**3**	**4**
adv	0.5	-0.044	-0.128	0.081
adv_time	0.033	-0.044	0.377	-0.234
advl_stance_all	0.314	-0.031	0.023	-0.055
conj_advl	0.425	0.086	-0.188	0.027
coord_conj_phrs	-0.043	0.32	-0.003	-0.059
det_nn_stance	-0.017	0.063	-0.07	0.396
fact_vb_other	-0.029	-0.071	0.154	0.448
jj_att_other	0.329	0.06	0.089	-0.117
jj_att_pred	0.324	0.025	0.018	-0.04
jj_attr	-0.042	0.285	-0.443	-0.195
jj_epist_pred	0.335	-0.134	-0.105	0.189
jj_pred	0.505	-0.337	-0.207	0.221
likely_vb_other	-0.16	-0.135	-0.124	0.313
mod_necess	0.397	0.092	0.047	-0.065
mod_poss	0.583	0.092	-0.134	-0.17
mod_pred	0.458	0.032	0.078	-0.09
nn_cog	-0.065	0.394	0.056	0.406
nn_common	-0.404	-0.296	-0.182	-0.073
nn_concrete	-0.05	-0.341	0.032	-0.219
nn_human	-0.084	0.329	0.071	0.123
nn_nom	-0.206	0.667	-0.163	0.109
nn_premod	-0.398	-0.025	-0.204	-0.113
nn_process	-0.039	0.369	-0.108	0.055
nn_quant	-0.136	-0.418	0.038	0.067

nn_stance_other	-0.033	0.203	-0.053	0.466
nn_stance_pp	-0.032	0.13	-0.134	0.462
nn_technical	-0.077	-0.275	0.131	-0.108
passive_short	-0.184	-0.224	0.419	-0.245
prep	-0.088	-0.379	-0.176	0.224
pro_3	0.108	0.417	0.379	0.175
pro_dem	0.459	-0.142	0.069	0.132
pro_it	0.563	-0.065	-0.008	-0.108
split_aux	0.171	0.006	0.028	-0.222
sub_conj_cond	0.305	0.006	0.182	-0.048
th_jj_stance_all	0.408	-0.163	0.003	-0.03
th_vb_stance_all	0.441	0.052	-0.023	0.15
that_rel	0.18	0.276	-0.043	-0.016
to_nn_stance_all	0.139	0.305	0.093	-0.023
to_vb_stance_all	0.37	0.232	0.279	-0.02
tt_ratio	0.042	0.429	0.153	-0.382
vb_act	0.033	-0.038	0.481	-0.096
vb_aspect	-0.03	0.086	0.345	-0.028
vb_be	0.618	-0.02	-0.058	-0.125
vb_comm	0.133	0.424	0.387	0.107
vb_mental	0.109	0.064	0.244	0.368
vb_past	-0.383	-0.144	0.539	0.13
vb_present	0.654	0.102	-0.261	-0.075
vb_progress	0.022	0.172	0.319	0.003
vb_suasive	0.044	0.377	0.128	-0.018
word_length	-0.317	0.797	-0.318	-0.048

Figure C1. Scree plot of the four-factor solution

Scree Plot
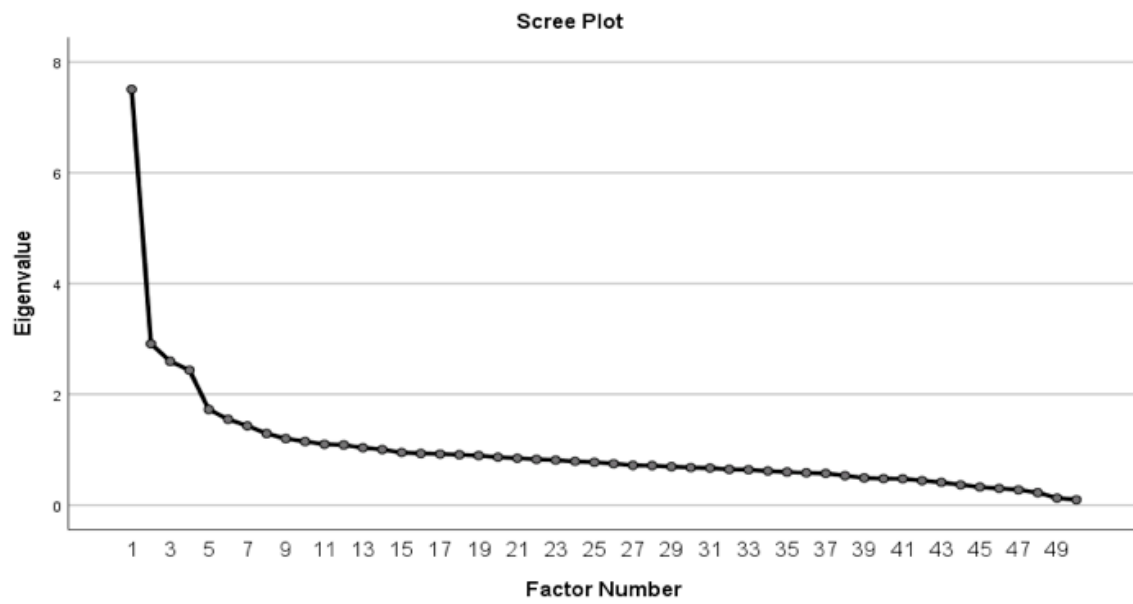
Post-hoc analysis: Dimension 1

Table C2
ANOVA (F = 316.94, p = <.0001, R² = .29)
Post-hoc comparisons (Games-Howell) for Dimension 1 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
introduction–discussion	-8.1	16.8	<.01	<.01
methods–discussion	-14.9	30.7	<.01	<.01
results–discussion	-9.3	18.6	<.01	<.01
methods–introduction	-6.8	15.2	<.01	<.01
results–introduction	-1.2	2.6	0.05	0.32
results–methods	5.6	12.1	<.01	<.01

Table C3
ANOVA (F = 262.68, p = <.0001, R² = .43)
Post-hoc comparisons (Games-Howell) for Dimension 1 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
hard introduction–hard discussion	-6.7711	10.982	<.01	<.01
hard methods–hard discussion	-15.6836	26.127	<.01	<.01
hard results–hard discussion	-8.8866	13.509	<.01	<.01
soft discussion–hard discussion	**7.3457**	**11.06**	**<.01**	**<.01**
soft introduction–hard discussion	-2.011	3.087	0.04	1.00
soft methods–hard discussion	-6.7644	11.3	<.01	<.01
soft results–hard discussion	-2.2889	3.605	0.01	0.23
hard methods–hard introduction	-8.9126	16.463	<.01	<.01
hard results–hard introduction	-2.1156	3.499	0.01	0.33
soft discussion–hard introduction	14.1167	23.089	<.01	<.01
soft introduction–hard introduction	**4.7601**	**7.965**	**<.01**	**<.01**
soft methods–hard introduction	0.0067	0.012	1.00	1.00
soft results–hard introduction	4.4822	7.734	<.01	<.01
hard results–hard methods	6.797	11.561	<.01	<.01
soft discussion–hard methods	23.0293	38.706	<.01	<.01
soft introduction–hard methods	13.6727	23.54	<.01	<.01
soft methods–hard methods	**8.9193**	**17.125**	**<.01**	**<.01**
soft results–hard methods	13.3948	23.827	<.01	<.01
soft discussion–hard results	16.2323	24.858	<.01	<.01
soft introduction–hard results	6.8757	10.741	<.01	<.01
soft methods–hard results	2.1223	3.62	0.01	0.22
soft results–hard results	**6.5978**	**10.586**	**<.01**	**<.01**
soft introduction–soft discussion	-9.3566	14.471	<.01	<.01
soft methods–soft discussion	-14.11	23.782	<.01	<.01
soft results–soft discussion	-9.6345	15.295	<.01	<.01
soft methods–soft introduction	-4.7534	8.208	<.01	<.01

	diff	t	p	p.adjusted
soft results–soft introduction	-0.2779	0.451	1.00	1.00
soft results–soft methods	4.4755	7.986	<.01	<.01

Table C4

ANOVA (F = 139.62, p = <.0001, R^2 = .30)

Post-hoc comparisons (Games-Howell) for Dimension 1 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
L1 introduction–L1 discussion	-9.52	13.38	<.01	<.01
L1 methods–L1 discussion	-15.85	22.6	<.01	<.01
L1 results–L1 discussion	-10.32	14.31	<.01	<.01
LX discussion–L1 discussion	**-3.09**	**4.3**	**<.01**	**0.01**
LX introduction–L1 discussion	-9.74	14.36	<.01	<.01
LX methods–L1 discussion	-17.01	24.52	<.01	<.01
LX results–L1 discussion	-11.31	15.85	<.01	<.01
L1 methods–L1 introduction	-6.33	9.67	<.01	<.01
L1 results–L1 introduction	-0.8	1.18	0.94	1.00
LX discussion–L1 introduction	6.43	9.54	<.01	<.01
LX introduction–L1 introduction	**-0.23**	**0.36**	**1.00**	**1.00**
LX methods–L1 introduction	-7.49	11.59	<.01	<.01
LX results–L1 introduction	-1.79	2.68	0.13	1.00
L1 results–L1 methods	5.53	8.31	<.01	<.01
LX discussion–L1 methods	12.76	19.24	<.01	<.01
LX introduction–L1 methods	6.1	9.86	<.01	<.01
LX methods–L1 methods	**-1.16**	**1.82**	**0.6**	**1.00**
LX results–L1 methods	4.54	6.91	<.01	<.01
LX discussion–L1 results	7.23	10.57	<.01	<.01
LX introduction–L1 results	0.57	0.89	0.99	1.00
LX methods–L1 results	-6.69	10.18	<.01	<.01
LX results–L1 results	**-0.99**	**1.46**	**0.83**	**1.00**
LX introduction–LX discussion	-6.66	10.42	<.01	<.01
LX methods–LX discussion	-13.92	21.25	<.01	<.01
LX results–LX discussion	-8.22	12.16	<.01	<.01
LX methods–LX introduction	-7.26	11.9	<.01	<.01
LX results–LX introduction	-1.56	2.47	0.21	1.00
LX results–LX methods	5.7	8.78	<.01	<.01

Post-hoc analysis: Dimension 2

Table C5
ANOVA (F = 367.51, p = <.0001, R² = .50)
Post-hoc comparisons (Games-Howell) for Dimension 2 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
hard introduction–hard discussion	2.41	7.40	<.01	<.01
hard methods–hard discussion	-2.09	6.40	<.01	<.01
hard results–hard discussion	-4.44	13.00	<.01	<.01
soft discussion–hard discussion	**8.26**	**23.80**	**<.01**	**<.01**
soft introduction–hard discussion	9.16	25.40	<.01	<.01
soft methods–hard discussion	5.35	15.40	<.01	<.01
soft results–hard discussion	4.41	10.50	<.01	<.01
hard methods–hard introduction	-4.50	14.10	<.01	<.01
hard results–hard introduction	-6.85	20.60	<.01	<.01
soft discussion–hard introduction	5.85	17.30	<.01	<.01
soft introduction–hard introduction	**6.75**	**19.10**	**<.01**	**<.01**
soft methods–hard introduction	2.95	8.70	<.01	<.01
soft results–hard introduction	2.01	4.90	<.01	<.01
hard results–hard methods	-2.35	7.00	<.01	<.01
soft discussion–hard methods	10.35	30.40	<.01	<.01
soft introduction–hard methods	11.25	31.70	<.01	<.01
soft methods–hard methods	**7.44**	**21.90**	**<.01**	**<.01**
soft results–hard methods	6.50	15.70	<.01	<.01
soft discussion–hard results	12.70	35.90	<.01	<.01
soft introduction–hard results	13.60	37.00	<.01	<.01
soft methods–hard results	9.79	27.70	<.01	<.01
soft results–hard results	**8.85**	**20.90**	**<.01**	**<.01**
soft introduction–soft discussion	0.90	2.40	0.24	1.00
soft methods–soft discussion	-2.91	8.10	<.01	<.01
soft results–soft discussion	-3.84	9.00	<.01	<.01
soft methods–soft introduction	-3.81	10.20	<.01	<.01
soft results–soft introduction	-4.75	10.80	<.01	<.01
soft results–soft methods	-0.94	2.20	0.36	1.00

Table C6

ANOVA (F = 48.37, p = <.0001, R^2 = .12)

Post-hoc comparisons (Games-Howell) for Dimension 2 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
L1 introduction–L1 discussion	1.43	3.20	0.03	0.86
L1 methods–L1 discussion	-2.76	6.10	<.01	<.01
L1 results–L1 discussion	-4.03	7.90	<.01	<.01
LX discussion–L1 discussion	-1.06	2.20	0.37	1.00
LX introduction–L1 discussion	0.79	1.70	0.67	1.00
LX methods–L1 discussion	-3.31	7.10	<.01	<.01
LX results–L1 discussion	-5.30	10.30	<.01	<.01
L1 methods–L1 introduction	-4.19	9.50	<.01	<.01
L1 results–L1 introduction	-5.47	10.90	<.01	<.01
LX discussion–L1 introduction	-2.49	5.20	<.01	<.01
LX introduction–L1 introduction	-0.64	1.40	0.84	1.00
LX methods–L1 introduction	-4.74	10.40	<.01	<.01
LX results–L1 introduction	-6.73	13.30	<.01	<.01
L1 results–L1 methods	-1.28	2.50	0.18	1.00
LX discussion–L1 methods	1.70	3.50	0.01	0.28
LX introduction–L1 methods	3.55	7.80	<.01	<.01
LX methods–L1 methods	-0.55	1.20	0.93	1.00
LX results–L1 methods	-2.54	5.00	<.01	<.01
LX discussion–L1 results	2.98	5.60	<.01	<.01
LX introduction–L1 results	4.82	9.50	<.01	<.01
LX methods–L1 results	0.72	1.40	0.85	1.00
LX results–L1 results	-1.27	2.30	0.31	1.00
LX introduction–LX discussion	1.85	3.80	<.01	0.11
LX methods–LX discussion	-2.25	4.60	<.01	<.01
LX results–LX discussion	-4.24	7.90	<.01	<.01
LX methods–LX introduction	-4.10	8.80	<.01	<.01
LX results–LX introduction	-6.09	11.80	<.01	<.01
LX results–LX methods	-1.99	3.80	<.01	0.10

Post-hoc analysis: Dimension 3

Table C7
ANOVA (F = 170.35, p = <.0001, R^2 = .42)
Post-hoc comparisons (Games-Howell) for Dimension 3 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
hard introduction–hard discussion	-0.54	2.60	0.15	1.00
hard methods–hard discussion	7.48	25.10	<.01	<.01
hard results–hard discussion	1.72	8.40	<.01	<.01
soft discussion–hard discussion	-0.25	1.30	0.90	1.00
soft introduction–hard discussion	-0.78	4.20	<.01	0.02
soft methods–hard discussion	3.80	15.30	<.01	<.01
soft results–hard discussion	1.40	6.10	<.01	<.01
hard methods–hard introduction	8.02	25.90	<.01	<.01
hard results–hard introduction	2.27	10.20	<.01	<.01
soft discussion–hard introduction	0.30	1.40	0.85	1.00
soft introduction–hard introduction	-0.23	1.10	0.95	1.00
soft methods–hard introduction	4.34	16.60	<.01	<.01
soft results–hard introduction	1.94	8.00	<.01	<.01
hard results–hard methods	-5.75	18.60	<.01	<.01
soft discussion–hard methods	-7.72	25.80	<.01	<.01
soft introduction–hard methods	-8.25	28.00	<.01	<.01
soft methods–hard methods	**-3.68**	**10.90**	**<.01**	**<.01**
soft results–hard methods	-6.08	18.70	<.01	<.01
soft discussion–hard results	-1.97	9.50	<.01	<.01
soft introduction–hard results	-2.50	12.50	<.01	<.01
soft methods–hard results	2.07	8.00	<.01	<.01
soft results–hard results	-0.33	1.30	0.88	1.00
soft introduction–soft discussion	-0.53	2.80	0.09	1.00
soft methods–soft discussion	4.04	16.20	<.01	<.01
soft results–soft discussion	1.65	7.20	<.01	<.01
soft methods–soft introduction	4.57	18.70	<.01	<.01
soft results–soft introduction	2.17	9.70	<.01	<.01
soft results–soft methods	-2.40	8.60	<.01	<.01

Table C8

ANOVA (F = 149.11, p = <.0001, R^2 = .37)

Post-hoc comparisons (Games-Howell) for Dimension 3 (p values Bonferroni corrected)

	diff	t	p	p.adjusted
L1 introduction–L1 discussion	-0.57	2.72	0.12	1.00
L1 methods–L1 discussion	5.63	19.02	<.01	<.01
L1 results–L1 discussion	1.76	7.34	<.01	<.01
LX discussion–L1 discussion	-0.55	2.91	0.07	1.00
LX introduction–L1 discussion	-1.06	5.30	<.01	<.01
LX methods–L1 discussion	5.28	17.44	<.01	<.01
LX results–L1 discussion	1.06	5.07	<.01	<.01
L1 methods–L1 introduction	6.19	20.90	<.01	<.01
L1 results–L1 introduction	2.32	9.68	<.01	<.01
LX discussion–L1 introduction	0.01	0.06	1.00	1.00
LX introduction–L1 introduction	-0.50	2.47	0.21	1.00
LX methods–L1 introduction	5.84	19.28	<.01	<.01
LX results–L1 introduction	1.62	7.76	<.01	<.01
L1 results–L1 methods	-3.87	12.12	<.01	<.01
LX discussion–L1 methods	-6.18	21.72	<.01	<.01
LX introduction–L1 methods	-6.69	22.95	<.01	<.01
LX methods–L1 methods	-0.35	0.95	0.98	1.00
LX results–L1 methods	-4.57	15.37	<.01	<.01
LX discussion–L1 results	-2.31	10.27	<.01	<.01
LX introduction–L1 results	-2.82	12.06	<.01	<.01
LX methods–L1 results	3.52	10.81	<.01	<.01
LX results–L1 results	-0.70	2.90	0.07	1.00
LX introduction–LX discussion	-0.51	2.77	0.11	1.00
LX methods–LX discussion	5.83	20.00	<.01	<.01
LX results–LX discussion	1.61	8.39	<.01	<.01
LX methods–LX introduction	6.34	21.25	<.01	<.01
LX results–LX introduction	2.12	10.48	<.01	<.01
LX results–LX methods	-4.22	13.88	<.01	<.01