A Corpus Analysis of Academic Writing and how it Informs Writing Instruction on a University Pre-Sessional Course.

A D Coates

PhD 2020

A Corpus Analysis of Academic Writing and how it Informs Writing Instruction on a University Pre-Sessional Course.

Andrew David Coates

Thesis submitted in fulfilment of the requirements of Manchester Metropolitan University for the degree of Doctor of Philosophy in Linguistics

Department of Languages, Information and Communications

Manchester Metropolitan University

2020

Abstract

Academic writing can pose difficulties to novice writers, and may be particularly challenging for second language (L2) English-speaking students. Pre-sessional English for academic purposes (EAP) courses aim to prepare L2 students for the linguistic demands they will face during their university courses; however, it is unclear the extent to which materials in such courses reflect any disciplinary variation in specific features of academic writing, or the issues that face learners in effectively using these features. In an effort to inform materials and pedagogy in the specific context of a UK higher education institution (HEI) pre-sessional course, a corpus of academic writing from eight disciplines was built, consisting of assignments by L2 and L1 Level 7 (master's level postgraduate) students, and published research articles. Four specific features of academic writing – academic vocabulary, lexical bundles, hedging, and citation - were then investigated, with a view to identifying how the various disciplines and writer groups may differ in the use of these features. This analysis was combined with a small qualitative element in the form of semi-structured interviews with L2 students who had previously participated in the pre-sessional course, and with an examination of the existing materials used for writing instruction on the course. The study identifies a number of areas in which recommendations could be made to amend and adapt writing instruction to more effectively address learner needs and acknowledge the importance of disciplinary variation. The recommendations not only address, for the first time, these four writing features with reference to the specific pre-sessional context, but also represent a useful model that can be applied to EAP writing instruction in HEIs more widely, and can potentially help to optimise teaching outcomes for L2 learners when it comes to the complexities of university English academic writing.

Acknowledgements

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Academic Writing

Academic writing is an important skill for under- and postgraduate university students, and while it may present problems for any novice writer, it can be particularly demanding for second language English speakers if they wish to participate successfully in their academic discourse communities (Bailey, 2014; Chang & Kuo, 2011; Gilquin et al., 2007). Writing in an academic context demands a consideration not only of generally accepted norms of the register, such as an avoidance of contractions and colloquialisms, and the use of appropriate vocabulary, but also of the specific conventions that may exist in individual disciplines (Yakhontova, 2006). These may include the use of specialised vocabulary, lexical bundles, modality and other hedging devices, passive constructions, attribution and citation, and writer stance (Hundt et al., 2016; Hyland, 1999b; Hyland & Tse, 2007; Liu, 2012). Academic writing, as Jalilifar (2012a:26) succinctly observes, 'requires sound knowledge of writing rubrics and principles underlying an academic text'.

Disciplinary variation has also been a well-recognised feature of academic writing for some time. Hyland (2002a) for example, contends that the research suggests there is a lack of uniformity in academic discourse, observing too that the writing tasks undertaken by university students vary according to discipline and educational level. He notes previous studies (such as Bridgeman & Carlson, 1984; and Casanave & Hubbard, 1992) that have demonstrated distinct differences between tasks given in humanities and social sciences compared to those in science and technology (with the former often involving analysis and synthesis of multiple sources, and the latter focusing more on describing procedures and conceiving solutions, for example) and between tasks in business studies and in engineering, as well as between those

8

in chemistry, maths and computer science. More recently, Taş (2010:122) argues that academic writing competence, rather than relying solely on linguistic proficiency, also depends on the 'awareness of rhetorical features of writing accepted by the discourse community', and indeed acceptance by one's disciplinary community depends upon learning its conventions (Flowerdew, 2000). Yakhontova (2006:154) makes the case that in any consideration of academic discourse, which 'exists in a variety of disciplinary realizations,' the 'influence of the professional subculture' must be taken into account. Not only does previous research confirm the notion that there is systematic variation in academic writing across disciplines (Li & Wharton, 2012), but the variation in discourse resulting from academic discipline can be stronger than that resulting from cultural background or first language (Ädel & Römer, 2012). As has been demonstrated in a number of studies (such as Kanoksilapatham, 2015; Jalilifar, 2012b; and Ozturk, 2007), variation can be not only inter-disciplinary, but also sub-disciplinary. Academic writing is, according to Klimova (2012:314), 'becoming increasingly more heterogeneous and blended in its discourses and mixing of modes'. Overall, this variation can be an obstacle in the path of those who may have difficulty understanding or accepting the particular practices of their academic field when it comes to academic writing competence (Hu, 2007).

The genres and text types of academic writing that students are required to produce at university level may also be quite different from those of which learners have previous experience. In a 2007 study, Hu observes that Chinese students attending an English for Academic Purposes (EAP) course designed to prepare them for undergraduate studies in Singapore, had experienced only relatively simple writing tasks, such as composing short letters or recounting personal experiences, and found expository writing challenging. Similarly,

international students wishing to study at UK universities are generally required to demonstrate their English language proficiency by means of securing a suitably high score on the International English Language Testing System (IELTS) test, and while this purports to be 'academic' in nature, the writing component of this test may bear little resemblance to what awaits second language English-speaking (L2) students once they arrive at university. Although Hawkey (2006), for example, contends that IELTS writing tasks are academic and meet the needs of prospective under- and postgraduates, other research has cast doubt on this assertion. Moore and Morton (2005) for example, draw two overarching conclusions from their study comparing IELTS and university writing in the Australian context – firstly that the writing tasks required of students at university vary hugely, suggesting that the rather limited scope of tasks in the IELTS test may not provide a great comparator, and secondly that there are indeed important differences between these two contexts in terms of their writing demands; while IELTS tasks tend to be spontaneous, opinion-based, anecdotally evidenced, concerned with real-world phenomena, and distinct from reading, the tasks students must complete at university are more likely to be non-spontaneous, requiring of evidence from research or scholarly sources, abstract, and inextricably linked to the reading process. This is not to mention the fact that the length of the writing tasks in the IELTS test is at most 250 words, compared to university assignments, which are generally of considerably greater length. In short, the language skills needed for higher education studies, and in particular those relating to academic writing, may not necessarily be acquired by means of sitting screening tests such as IELTS (Green, 2007).

Even those learners who may have previous experience of writing more akin to that undertaken in university contexts, may still be challenged by the diversity in tasks, genres and

disciplines they come to face. This is an issue that can pose challenges even to first language English-speaking (L1) students. As far back as the mid-nineties, Lea and Street (1998:164) found students in two UK universities citing problems in coping with the demands of writing not only for different subjects, but also for different lecturers, one participant noting that 'Everybody seems to want something different' and prompting the researchers to conclude that the concept of transferable, generic skills across subjects, disciplines and departments was perhaps questionable. This notion is further explored by Hyland (2006:20), who describes disciplines as 'sites where differences in worldview or language usage intersect as a result of the myriad backgrounds and overlapping memberships of participants.' In their 2013 study, Hardy & Römer analysed A-graded writing from the Michigan Corpus of Upper-level Student Papers (MICUSP) and found that disciplines varied widely in terms of whether writing could be categorized as 'narrative', such as in philosophy or education, or 'descriptive and informational', such as in biology or physics. They also found a dichotomy between 'non-procedural evaluation', and 'procedural discourse', the former being characterised by adverbials, while the latter relies more heavily on nouns and passives (highlighting the difference between, for example, English and sciences). L2 students may find the discipline-specific conventions of the academic discourse community they have joined difficult to adjust to (Granger & Paquot, 2010; Hyland, 2008a, 2008b), and this presents them with another potential obstacle in addition to those they might face from attempting to write in a language that is not their first, and from any general unfamiliarity with the Western notion of what constitutes academic discourse.

All of this means that L2 students potentially face a number of challenges when it comes to academic study, particularly writing, at university level, and higher education institutions

therefore require, in addition to general academic achievement, some evidence of language proficiency. As alluded to above, L2 international students wishing to attend UK universities are required to demonstrate their English proficiency by means of the IELTS test. However, applicants who are unable to attain the required score may still be accepted on condition that they undertake a period of additional academic English study, often run in the summer months prior to the start of academic courses in the autumn term, and known as a pre-sessional course.

1.2 EAP Pre-sessional Courses

Pre-sessional English courses aim to assist international L2 students to reach a level of ability in the language that will enable them to successfully complete their programs of study, and a significant component of this process is learning the features, characteristics, and techniques of academic writing. However, there may be some cause to question how accurately the content of such courses reflects the reality of how academic writers write. As Harwood (2005:150,153) notes for example, many EAP textbooks 'understate the enormous disciplinary variation in style and language which corpora reveal…Studies which have compared what EAP textbooks teach with corpora of expert and/or student academic writing find textbooks wanting.' Additionally, in many cases pre-sessional students study in mixed groups, rather than being grouped by academic discipline (Thompson, 2006), such that an individual whose future degree programme is biology for example, might find themselves learning academic English next to prospective law or engineering students. Given the disciplinary variation within academic discourse identified in the literature, and the fact that different disciplines have different expectations when it comes to the features and forms of academic writing (Paltridge, 2002), it may be difficult, in this situation, to optimise the pedagogical outcomes of pre-sessional courses when it comes to academic writing, and to ensure that course content

accurately reflects the reality of this discourse genre and goes some way to encompassing the variation found across disciplines.

1.3 Corpus Linguistics

One way to usefully inform pre-sessional course content with observations taken directly from academic writing itself, rather than from the intuitions of materials writers (Gilquin et al., 2007) is to employ a corpus linguistics approach. This argument for the value of utilising corpora in informing teaching content is reiterated by Aull (2015), who contends that the systemic analysis of language patterns in academic writing is an under-utilised tool in guiding students to be more aware of such features. If one hopes to identify the linguistic characteristics of a particular discourse type, examples of it are required. Corpus studies, by analysing such examples, have provided a variety of useful insights into the features of academic discourse, and have also shed light on areas such as disciplinary variation (Thompson, 2006). Prior to the development of corpus analysis software packages, building an accurate picture of how language is used in academic discourse was extremely challenging, but corpus research now offers significant opportunities in this regard (Gilquin et al., 2007), and with an understanding of the conventions and characteristics of academic writing, comes the opportunity to apply this knowledge to pedagogy. Indeed, Wingate & Tribble (2011:7) go so far as to say that the construction and analysis of academic corpora can lead to 'a dramatic change in the specificity of the writing instruction that can be offered to EAP learners across a wide range of disciplines'.

For corpus-based research to be of optimal benefit, it is vital to consider how academic corpora are constructed and utilised. The details of this clearly depend on the specifics of the research itself, and what it aims to achieve, but in the context of potentially informing EAP course content, the sources used to compile the corpora to be analysed are an important

consideration. Nesi et al. (2004) note that while expert (published) writing can be a very useful source of study, it is important to bear in mind that it has a different readership and purpose than student writing. They suggest that a valuable approach is to examine the writing of L1 students during their courses of study. Much more is known about, for example, published research articles than about the writing of novice, student academic writers (Ädel and Römer, 2012), which also suggests that analysing student writing may prove to be a useful practice.

 If the purpose of a corpus study is to highlight areas of need in an EAP context however, it is clear that simply looking at one source, be that expert writing or the work of students, will not prove sufficient; some manner of comparison is required. Gilquin et al (2007) tackle the somewhat contentious issue of what form such a comparison might take. They observe that the notion of comparing expert and L2 student writing has not been without its critics, and that some researchers have suggested a more useful comparison might be made by examining writing from L1 and L2 students. They go on to argue however, that in the context of their own study, which sought to assist learners with their writing skill, an expert corpus was judged to be a better target model, in part because L1 students may not always produce writing that L2 students would wish to imitate. The notion of comparing L2 student and expert writing is one also espoused by Ädel (2006) and by Bolton et al. (2002), who contend that published writing represents the 'target' for both L1 and L2 student writers. Perhaps the most comprehensive picture of academic discourse, particularly when it comes to identifying needs and informing EAP course content, would be one built up by conducting a three-way analysis, so as to include published 'expert' writing, and that of both L1 and L2 students.

 Corpus analysis has been employed as a teaching tool in various EAP studies. For example, Charles (2012) examined the attitudes of L2 PhD and master's students building their own

14

discipline-specific corpora. Yoon (2008) investigated the ways in which EAP students' writing process is affected by utilising a corpus during their classes, and a number of other studies have also looked into this kind of direct corpora use in pedagogy (See Charles, 2007; Charles, 2011; Varley, 2009; Yoon & Hirvela, 2004). As Braun (2005) notes, corpora have been employed for direct use by both teachers and learners in a variety of ways. However, their potential as a basis for informing pedagogy in the context of EAP in higher education institutions has been less widely explored. The present study is carried out on the principle that when considering academic writing pedagogy, and designing course materials, analysis of academic corpora may prove to be a beneficial methodology, one which could potentially lead to more successful outcomes for international students in UK higher education institutions.

 While corpus analysis represents a potentially useful means by which to inform course materials for EAP writing, needs analysis often benefits from the use of multiple sources and the perspective of the 'insider' (Jasso-Aguilar, 1999). To rely solely on 'data-driven "objective" information about learners' without also considering their self-knowledge and the more subjective insights they may be able to provide, is a questionable methodology (Belcher, 2006:136). As Long (2005:20) observes, 'learners sometimes not only wish to be consulted, but also are well informed', and the notion that needs analysis is best conducted as an ongoing process involving learners is one also highlighted by Abdullah (2005). In a study of L2 graduate students at a Canadian University, the students' perceived language difficulties were found to be closely linked to their academic performance (Berman & Cheng, 2010), and thus it would seem sensible to take these perceptions, along with other insights the learners themselves may be able to contribute, into account when attempting to inform EAP course content. Combining subjective data from students with objective data from corpus analysis provides a

degree of triangulation, which is valuable in that it reduces the likelihood of systemic bias and increases validity (Dörnyei, 2007; Long, 2005).

It is also important however, to consider the extent to which learners constitute a potentially reliable source of information. Individuals who are 'pre-experience' (in this context, international students who have yet to either participate in a pre-sessional EAP course or have at least some experience of their academic courses proper) may not be appropriately situated to provide useful insights when it comes to the challenges of academic writing or the benefits of their EAP instruction. An example of this is the study by Beatty and Chan (1984), in which perceived academic needs were found to differ markedly between students who were at the point of entering an English-medium study environment, and those who had been in one for at least six months. It is therefore important that learners be involved at an appropriate stage in their studies, so as to gain optimal benefit from their insights, and this should, as previously alluded to, be combined with other sources of data (Long, 2005).

Overall then, academic writing plays an intrinsic role within higher education courses, particularly at postgraduate level, and it is vital that students become familiar with the features and conventions specific not only to this discourse genre in general, but also to the academic discipline communities within which they will study. Compared to their L1 counterparts, L2 international students can face particular challenges in becoming proficient academic writers, and it is therefore important that when they undertake pre-sessional EAP courses, these courses provide optimal learning and study opportunities when it comes to academic writing instruction. A mixed methods research approach, combining the corpus analysis of academic writing by L1 students, L2 students, and published sources with qualitative data from interviews conducted with L2 students themselves will establish the potential for corpus

linguistics to usefully inform teaching materials in the writing component of a university pre-sessional EAP course, and assist in optimising the learning outcomes for students participating in these courses.

The context chosen for the study is a university in the North-West of England. The institution has a large student population, with 33,050 students in the 2018/19 academic year (ranked 11th in the UK in terms of student numbers), of which 6735 were postgraduates. There is also a sizeable and diverse population of international students at the institution (2600 non-UK domicile students in the 2018/19 academic year) This makes the university an ideal context in which to situate a study such as this (all figures sourced from HESA, 2020).

Four aspects of academic writing have been chosen as the focus of the analysis, namely i) academic vocabulary, ii) lexical bundles, iii) hedging, and iv) citation (sections 2.1.3(i) – 2.1.3(iv) provide full definitions and in-depth analysis of all four terms). The selection of these four features was made with both the limitations of IELTS testing (in terms of comparability to the demands of academic writing at university) and the importance attributed to these features in the literature in mind. While lexical resource is a consideration in IELTS writing skills, the extent to which this includes what might be considered 'academic vocabulary' is unclear, particularly in terms of specialised, subject-specific vocabulary. Similarly, IELTS writing does not give any real attention to the use of lexical bundles that may be appropriate in the academic register of varying disciplines. Bundles (in other words, extended collocations), rather than shorter, two-word collocations, have been chosen for a number of reasons. Firstly, as Hyland (2008a:5) points out, bundles play a role in giving a sense of distinctiveness to a register:

> Thus the presence of extended collocations like *as a result of*, *it should be noted that*, and *as can be seen* help identify a text as belonging to an academic register while *with regard to*, *in pursuance of*, and *in accordance with* are likely to mark out a legal text.

There are thus clear advantages to identifying the bundles associated with different academic disciplines if this allows student writers to gain communicative competence in their field. Secondly, there exists a framework for the understanding of how bundles are used in academic discourse (see Hyland, 2008b and section 2.1.3 (ii) for details). The existence of such a framework facilitates a more in-depth analysis of lexical bundles in academic writing. Finally, while L2 students may have received some instruction in shorter and more general collocations (even if this is limited), general English courses and even 'academic' IELTS instruction are unlikely to have focused on the extended lexical bundles and their disciplinary specificity, which makes them an important element in higher education EAP. IELTS instruction, similarly, does not give a great deal of attention to hedging, and citation is completely beyond its remit.

These features have also all been demonstrated to play important roles in academic discourse, display significant disciplinary variation, and present potential challenges to novice, and in particular, L2 academic writers (Biber et al., 1999; Charles, 2006; Friginal et al., 2014; Gilquin et al., 2007; Hinkel, 2009; Hyland, 1999a, 1999b, 2008a, 2008b; Hyland and Milton, 1997; Leki & Carson, 1997; Liu, 2012; Petrić, 2007; Woodward-Kron, 2008 to name just some of the studies) and therefore represent important elements in the teaching of academic writing at university level. Citation use and hedging for example, have been associated with higher scores in academic writing (Ferris, 1994; Petrić, 2007). It may be true that these features are not the primary impediments to academic success, but they are important aspects of academic discourse, and unlike features such as the passive voice, relative clauses, or cohesive markers for example, are not likely to have been routinely covered in IELTS or general English

instruction. They are therefore of greater potential importance in the context of pre-sessional courses.

1.4 Research Questions

The study will seek to assess how far a three-way corpus analysis of published, postgraduate L2, and postgraduate L1 academic writing in a variety of disciplines, combined with qualitative semi-structured interviews with pre-sessional course participants, and an examination of existing course materials, can usefully inform pedagogy for the writing component of a pre-sessional course at a university in the North-West of England. In order to do this, the study will address the following research questions:

RQ1. What is the extent of disciplinary variation in the use of academic vocabulary, lexical bundles, hedging and citations?

RQ2. How do L2 student, L1 student, and published academic writing differ in the usage of academic vocabulary, lexical bundles, hedging and citations?

RQ3. How are academic vocabulary, lexical bundles, hedging and citations represented in existing pre-sessional writing materials?

RQ4. What are the perceptions of students who have completed the pre-sessional course in terms of the teaching of these four academic writing features, and their own difficulties with academic writing?

The answers to the research questions will provide a comprehensive picture of academic writing across the disciplines examined, with reference to the similarities and differences between L1 and L2 student, and published writing. This will make clear the extent to which the institution's pre-sessional course can be usefully informed by the corpus analysis, and

suggestions will be made to adapt the materials in line with the findings, such that other institutions may apply these principles to their own courses and materials design, extending the benefits of the proposed research from the specific context in which it is carried out, to the wider higher education community as a whole. The student corpus itself will also constitute a valuable tool for future research. If training international L2 students in academic writing can be made more effective, this can only help to lead to greater academic success – an optimal outcome for institutions and students.

(A note on terminology – the terms L1/L2 are used here so as to avoid the 'native speaker' (NS) and 'non-native speaker' (NNS) labels, which, while they are still commonly encountered in the literature, are also potentially problematic, raising issues related to the notions of native-speakerism, ethnocentrism, and the idea that this classification automatically suggests NS superiority, or some lack on the part of NNSs. It should also be made clear that while it may be a false dichotomy to categorise English users on the basis of whether English is or isn't their first language, the practicalities of the study are such that it is not possible to establish the particular language backgrounds or proficiencies of all those individuals contributing to the corpus, and as such, some manner of relatively straightforward labelling system is required to discriminate between English users – even if that system is not ideal.

The term 'international student' is used here solely to distinguish between UK domicile students, and those whose permanent residence lies outside the UK.)

2 Review of the Literature

2.1 Overview of Chapter

This chapter reviews the findings of previous studies in order to clarify and examine in depth, the context and background of this study, and to further situate it within the existing knowledge base. The literature is considered in three broad sections: academic writing, EAP, and corpus linguistics.

2.2 Academic Writing

2.2.1 Overview of Academic Writing

Academic writing is a vital skill for higher education students, who must familiarise themselves with the knowledge and strategies demanded in this genre (Hu, 2007). This is particularly true if they wish to successfully participate in the academic discourse community beyond undergraduate level (Flowerdew, 2000), as dissertations and theses represent formidable challenges, in part due to their length, but also as a result of the high standards to which they are held (Dong, 1998). Nonetheless, developing competence in academic writing can be demanding (Hu, 2007). It is a sophisticated and complex genre, presenting difficulties to both L1 novice and L2 writers, who may be unfamiliar with the appropriate style and register, one which, as Chang and Kuo (2011:223) observe, manifests 'at various linguistic levels, including lexico-grammatical features, rhetorical functions, writing skills, and generic structures'. Academic genres require writers to anticipate the background knowledge of their audience, develop argumentation strategies, and control their use of hedging (defined by Hyland (1996:433) as 'the expression of tentativeness and possibility…central to academic writing, where the need to present unproven propositions with caution and precision is essential'). In addition, they must employ effective metadiscourse and signposting, utilize appropriate multi-

word expressions (lexical bundles) and vocabulary, and successfully incorporate, cite, summarise, and paraphrase source material. All of this must be done while conforming to the many conventions of academic discourse, which can vary widely across disciplines (Hirvela & Du, 2013; Hyland, 1994, 2004, 2008a; Paltridge, 2002). These challenges may be more pronounced for students who are still learners of the language in which they must write (Ädel & Erman, 2012).

2.2.2 Overview of L2 Issues in Academic Writing

Studies examining L2 student academic writing in English have highlighted a variety of areas in which these students face increased difficulties compared to L1 students. Dong (1998) notes issues including writing concisely and correctly, and difficulty with specific knowledge related to discipline, genre, and audience, while Hyland (2008b) finds L2s facing challenges with multi-word clusters. Additionally, language for hedging may be used with less control by L2 English learners (Chen & Baker, 2010). Gilquin et al. (2007) observe that a number of studies have shown a general problem with learners employing a style more akin to that of speech in academic writing, including the overuse of first and second person pronouns, and the underuse of features associated with formal writing, such as EAP words like *argument, advocate*, and *issue*. They also highlight further problems raised by analysis of L2 student corpora, including general writing issues such as semantics, phraseology, and positioning, as well as specific difficulties posed by the academic register itself.

Cultural variations may also play a role in determining the success with which L2 students can adapt to the Western model of academic writing. Steinman (2003:82) for example, notes that while in English 'the onus is on the writer to make things clear…the Japanese are more likely to expect...the reader to make sense of the text,' and that presentation and argumentative

style may also vary depending on the cultural background of the writer. Hirvela & Du (2013) also draw attention to this issue of cultural background, arguing that L2 writers transitioning between languages and rhetorical systems may find processes such as paraphrasing, summarizing, and quoting particularly difficult.

Nor are the problems encountered by L2s writing academically in English necessarily confined to the experiences of under- and postgraduate students. In a study of L1 Chinese university lecturers and professors in Hong Kong, Flowerdew (1999) notes that concerns raised by the participants in terms of the disadvantages they face when attempting to publish their work in journals that require English, include lack of vocabulary, difficulty asserting claims with appropriate force, and lacking the means to adequately express themselves.

It may be helpful at this point, to consider in more depth some of the aforementioned and most commonly observed features of academic writing, upon which this study will focus, and the associated problems experienced by L2 writers.

2.2.3 Specific Features of Academic Writing and the Issues Faced by L2s.

2.2.3.1 Vocabulary

It is important, first of all, to examine what *academic vocabulary* really means. The notion of an academic vocabulary centres on the idea that certain items are 'reasonably frequent in a wide range of academic genres but are relatively uncommon in other kinds of texts' (Hyland and Tse, 2007:235). A great deal of research has been conducted in an effort to identify these items and thus produce a list of 'academic vocabulary'. Early efforts, utilising corpora of university textbooks and lecture materials across a number of disciplines, or students' annotations of vocabulary items in their textbooks, include Campion & Elley (1971), Ghadessy

(1979), Lynn (1973), and Praninskas (1972). Subsequently, these lists were combined to form the UWL, or University Word List, comprising around 800 items occurring widely and frequently in academic texts, but not among the 2000 most common words in the GSL, or General Service List, compiled by West in 1953 (Xue & Nation, 1984). The culmination of all these efforts was perhaps the Academic Word List (AWL) produced by Coxhead (2000). This list of 570 word families (base words together with inflections and derivations) and 3107 individual items was based on a 3.5 million word corpus ranging across science, art, law, and commerce, with seven individual subject areas in each discipline.

Coxhead (2000:226) described the AWL as 'a specialised vocabulary with good coverage of academic texts, regardless of the subject area', which perhaps suggests that it represents a general academic vocabulary that would be of use to all students, no matter the discipline in which their studies are based. Nonetheless, this notion of universality has not been without its critics. Paquot (2007) notes that the AWL focuses on vocabulary necessary for academic reading comprehension, and that students' productive needs are quite different, and proposes the creation of a productively-oriented word list. Hyland and Tse's 2007 paper, which re-evaluates the AWL against a corpus of book reviews, textbook chapters, research articles, scientific letters, theses, and dissertations across eight disciplines, goes further, by questioning whether a general academic vocabulary exists at all. While accepting that certain words occur with greater frequency in academic texts, Hyland and Tse contend that 'It is by no means certain that there is a single literacy which university students need to acquire to participate in academic environments' (p.236). A number of reasons are put forward for this, including varying levels of coverage provided by the AWL in different disciplines, unequal distribution across the sub-corpora of both the most, and least frequently occurring words, and the fact

that words may take on different or specific meanings in different disciplines, and be associated with different collocates. They note, for example, that in social sciences, the word *process* is less likely to be encountered in noun form than it is by engineering and science students, and that the word *strategy,* while occurring with a generally high frequency across disciplines, tends to associate with the words *marketing* in business, *learning* in applied linguistics, and *coping* in sociology. Hyland and Tse (2007) conclude by saying:

> In sum, although the generic label academic vocabulary may be a convenient shorthand for describing a general variety, it conceals a wealth of discursive variability which can misrepresent academic literacy as a uniform practice and mislead learners into believing that there is a single collection of words which they can learn and transfer across fields. (p.251)

Coxhead herself, re-examining the AWL ten years after its creation, acknowledges that there is a need for work involving more balanced corpora that include a wider variety of subject areas (2011). Indeed, some work has been done more recently to compile lists of academic vocabulary drawn from larger corpora and including a wider range of academic disciplines (for example, Browne et al, 2013; Gardner & Davies, 2014). Many disciplines utilise specialist, subject-specific vocabulary, and it is vital that students are aware of this vocabulary and can use it accurately if they are to demonstrate their understanding of the subject matter (Woodward-Kron, 2008). The fact that various studies have been conducted to examine discipline or subject-specific academic vocabulary (e.g. Chen and Ge, 2007; Mudraya, 2006; Wang et al., 2008; Ward, 2009) also suggests that the idea of a general academic vocabulary has become a less useful one. Nonetheless, while they may not encompass the full range of variation that exists across disciplines in academic writing, lists such as the AWL do provide a useful starting point when it comes to delineating between vocabulary that might be considered appropriate for academic register, and that which may not.

It is perhaps not surprising, given this apparent disciplinary variation, that L2 students encounter a number of difficulties when it comes to their use of vocabulary in written academic discourse. Issues with L2 students' use of vocabulary in academic writing have been highlighted through the literature for many years. Leki & Carson (1997), in a study of the writing experiences of L2 students at an American university, found that the majority of their respondents noted difficulties in finding the vocabulary they needed to express their ideas when asked to write within a time limit. Flowerdew (1999) notes a perception among Hong Kong Chinese academics that they have a less rich vocabulary compared to their L1 English counterparts, while Dong (1998) observes that L2 students were 30% more likely to cite vocabulary-related problems than were L1 students when it came to writing theses or dissertations.

Additionally, L2 students may have issues with the overuse or underuse of vocabulary items. Learner corpora studies have shown that certain general, vague words such as *thing*, *people*, and *problem* are overused by L2 students, whereas other items regarded as appropriate for academic register, such as *argument*, *issue*, and *advocate* are underused (Gilquin et al., 2007). The same study also highlights issues of misuse, such as in the case of *on the contrary*, which learners have a tendency to employ simply to illustrate difference, rather than to show direct contradiction as L1 writers usually do. Gilquin & Paquot (2007), in an analysis of the International Corpus of Learner English (ICLE – Granger et al., 2009), which comprises around 6000 essays by L2 writers from 16 different L1 backgrounds, observe the use of various features that are more commonly associated with speech than with the formal academic register, such as emphasisers like *really* and *absolutely*. They suggest that this may be the result of a focus on grammatical accuracy rather than stylistic appropriacy, and of textbooks that pay

little attention to register, and present items as being completely interchangeable (*so* and *therefore*, for example). Bolton et al. (2002) examining the International Corpus of English in Hong Kong (ICE-HK), find similar underuse, overuse, and misuse among L2 writers in comparison to L1s when it comes to connectors such as *lastly*, *moreover*, *firstly*, *consequently*, *on the other hand*, *furthermore*, and *besides*.

Perhaps the area of academic vocabulary use that has received the most attention however, is that of multi-word sequences, or lexical bundles. Paquot (2005) observes that research has shown learners to overuse certain common collocations, while underusing others, notably what could be termed 'EAP multi-word sequences'. However, before examining learner issues in this area, it may be worth considering just how important these lexical bundles are in academic discourse.

2.2.3.2 Lexical Bundles

Lexical bundles, also referred to as chunks or clusters, are defined by Hyland (2008b:42) as 'words which follow each other more frequently than expected by chance'. Cortes (2004: 400) describes 'extended collocations, sequences of three or more words that statistically co-occur in a register' (*as a result of*, *the fact that the*, *in relation to the*, and *it should be noted that* would be examples in the academic register). Bundles play an important role in proficient language use, and are thus one element that contributes to competent academic writing (Cortes, 2004; Hyland, 2008a, 2008b; Liu, 2012). Wray and Perkins (2000) suggest that these bundles function as short cuts when processing language; in effect the bundle is processed holistically rather than sequentially. They go on to theorise that this decrease in the effort required in processing may go some way to explaining why a given discourse community will adopt and prefer certain bundles, even though there may be many other possible and

permissible combinations that could express the same idea. In this way, the use of these sequences plays a part in signalling the register (Hyland, 2008b), and allows academic writers to exhibit appropriate formality, and to signpost discourse stages (Peters & Pauwels, 2015), while meeting the expectations of the academic readership (Coxhead & Byrd, 2007). Conversely, a lack of appropriate bundles may be indicative of a writer who is a novice in a given discourse community (Hyland, 2008b).

As in the case of academic vocabulary in general, much effort has gone into identifying lexical bundles in academic writing, assessing their importance, and establishing the extent to which their use is affected by disciplinary variation. Biber et al. (1999), examining academic prose, found 4-word and 3-word bundles in general to occur frequently, and some individual bundles appeared over 200 times per million words. Similarly, Hyland (2008b) notes that some bundles found in a 3.5 million word academic corpus comprising dissertations, theses, and research articles, were extremely common. Peters and Pauwels (2015:29) likewise refer to the 'importance and omnipresence' of these sequences in academic writing. Arguing the value of an 'empirically devised and pedagogically useful' list of bundles, comparable to the AWL for academic vocabulary, Simpson-Vlach & Ellis (2010:487) created the Academic Formulas List (AFL). This combined measures of both frequency and formulaicity to identify over 200 core 3, 4, and 5-word bundles from a combination of spoken and written academic corpora.

While efforts have been made to ascertain how commonly lexical bundles occur in academic writing, other research has sought to understand how they are used in discourse. Biber et al. (2004:384) produced a taxonomy for lexical bundles in a university context, including textbooks, classroom teaching, conversation, and academic prose. The study classified lexical bundles into three main classes: 'Stance Expressions', such as *the fact that the* and *it is*

necessary to; 'Discourse Organisers', like *on the other hand*; and 'Referential Expressions', including *in the case of* and *on the basis of*. Hyland (2008b:49) went further by examining lexical bundle use in the aforementioned 3.5 million word corpus, which included texts from applied linguistics, microbiology, business studies, and electrical engineering, so as to focus specifically on academic writing. His taxonomy also consisted of three broad categories: 'Research-Oriented', with functions such as indicating time or place (*in the present study*), describing procedures (*the purpose of the*), or providing quantification (*a wide range of*); 'Text-Oriented', involved in organising the text by way of transitions and structure (*in contrast to the, in the next section),* and results and framing *(it was found that, on the basis of*); and 'Participant-Oriented', to illustrate the writer's evaluations (*it is possible that*) or directly address the reader (*it should be noted that*).

Identifying and categorising bundles may be useful in research, but this information can only be applied optimally to the teaching of academic writing, and to the potential challenges faced by L2 writers, if it is considered alongside an appreciation of the disciplinary variation in their use. Hyland (2008a) utilised the same 3.5 million word academic corpus to examine this variation, concluding that there are indeed notable differences between disciplines when it comes to lexical bundles. Biology and electrical engineering showed some commonality, as did business studies and applied linguistics, however when the bundles employed in the sciences were compared to those used in the social sciences, there was considerable contrast. Research-oriented bundles for example, were far more prevalent in science and engineering than in business studies or applied linguistics, where text-oriented sequences were dominant. Liu's (2012) study, at the time employing the two largest corpora yet used to investigate multi-word constructions (including lexical bundles) in academic writing, aimed to establish the most

frequently used sequences, and also found variation between occurrences in humanities, social sciences and law, and those in medicine, engineering and natural sciences. Cortes (2004) compared published and student writing in biology and history, and again discovered disciplinary variation. This came in the form of structure – in history, lexical bundles were all either noun phrases or prepositional phrases, while in biology the structures were much more varied, and in function – biology utilised hedging bundles with greater frequency than did history, and referential markers of time were employed in different ways. Opinions differ however on the extent of disciplinary variation in bundles, and thus the extent to which pedagogy need take a differentiated approach if they are to be taught. While Hyland (2008a) argues that the approach must necessarily be discipline specific since there are too few bundles common across multiple disciplines, Simpson-Vlach & Ellis (2010), in their aforementioned AFL, did produce a core list that they argue transcends disciplinary boundaries.

What issues then, do L2 writers encounter when it comes to the utilisation of lexical bundles in academic writing? The acquisition of formulaic sequences can be problematic for language learners, and in the context of academic writing, being sensitive to the discourse community's preference for one sequence over another can be very difficult, which can lead to great challenges for L2 writers attempting to produce texts that will be regarded as assured and authoritative by their readership (Hyland 2008a, 2008b). As Ädel and Erman (2012:81) also note, 'It is notoriously difficult to achieve idiomaticity, that is, the knowledge of conventionalized combinations of words, in academic discourse', an observation they apply to both L1 and L2 writers. That said, the same study reports that the variety of lexical bundles employed by L2 student writers was more restricted than in the case of L1s; 55 bundles were shared by both groups, and 60 were unique to L2 students, but 130 were used only by L1

students. Similar findings emerge from the study by Gilquin et al. (2007), who observe that, having a more limited selection of expressions at their disposal, L2s tend to overuse some of the ones they know. Durrant and Schmitt (2009) report differing, but in some ways complementary results, in that while L2 writers may use just as many collocations as L1s, they tend to overuse those that generally occur with high frequency, while underusing more novel, low-frequency expressions. Bychkovska & Lee (2017) found L2 undergraduate writers to use more bundles (both types and tokens) than L1 writers, although this result may be related to the fact that L1 disciplinary writing was compared to L2 writing from a general composition class, suggesting that these L1 writers may have been utilising bundles within the boundaries of their own disciplines, which may have limited their usage.

Overuse of the same bundles can cause writing to appear repetitive, and give the impression that the writer's lexical resources are more limited than they may actually be. An issue related to the underuse of appropriate lexical bundles is that this can lead to problems in register. As mentioned, the particular bundles preferred in academic writing act as indicators of that register; a lack of these may lead to writing appearing too informal or colloquial (Gilquin & Paquot, 2008; Peters & Pauwels, 2015). Cortes (2004) suggests that these issues may arise from L2 students never having been instructed in the use of the appropriate bundles, since merely encountering them in academic reading may not result in their acquisition for productive use. L1 students, on the other hand, are likely to have had much more extensive reading exposure, and therefore may not be in such need of instruction. The ability to successfully manipulate and utilise lexical bundles is clearly an important one for academic writers, assisting them as it does to conform to the expectations of their discourse community. Bundles play a role in indicating register, are important elements in a variety of rhetorical

functions, and can be markers of metadiscourse. Another important aspect of this, particularly in academic writing, is the notion of hedging.

2.2.3.3 Hedging

'The terms hedges and hedging generally refer to a large class of lexical and syntactic features of text that have the goal of modifying and mitigating a proposition' (Hinkel, 2005:29). Crompton (1997:281) defines the term as follows, 'A hedge is an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters.' Hedging, in other words, is a means by which a writer may express the possibility or uncertainty of an argument or inference, and is of great importance in academic writing, where it is necessary to 'present unproven propositions with caution and precision' (Hyland, 1996:433). Statements made by academic writers may need to convey varying degrees of uncertainty and confidence, and must also indicate appropriate modesty and deference to the audience (Hyland & Milton, 1997), while protecting the writer from false interpretations (Hyland, 2008b). As Poos and Simpson (2002: 4) state, 'English has a wealth of lexical resources for expressing uncertainty, lack of commitment to a proposition, and vagueness'. These include the use of modals, such as *can*, *might*, *could* and *may* (Hinkel, 2009); adjectives, such as *possible*, *likely*, and *conceivable*; adverbs, including *possibly* and *probably;* nouns, like *doubt* and *likelihood*; and verbs, such as *suggest* and *seem* (Hyland & Milton, 1997).

The use of hedging devices in academic writing is highly conventionalised, and particular to the specific discourse community (Hinkel, 2005). Failure to hedge claims and predictions appropriately can cause pragmatic issues as well as impacting negatively on register. Friginal et al. (2014) note that a higher frequency of hedges was one feature that predicted higher scores for L2s in studies of Test of Written English essay corpora, but both L1 and L2 academic

writers need to learn to strike the correct balance between vagueness and precision, and how to be 'appropriately imprecise' (Hinkel, 2009:672).

Like other features of academic writing, the use of hedging has been studied with regard to disciplinary variation. For example, Hyland (1999a) found that in an analysis of research articles from eight different disciplines, while hedges were the most frequent marker of writer stance across all the disciplines, they occurred more frequently in biology than in philosophy. Vázquez & Giner (2008) however, observe that hedging items were more common in marketing research articles than in either biology or mechanical engineering, although this was based on an analysis of only 12 research articles, compared to the 56 examined in Hyland's study. They posit that this may be due to the fact that marketing is a subject in which results are more open to interpretation, and are contextually influenced, whereas in 'harder' sciences, there is often more precise, numerical data, and thus less necessity to hedge. These results, and this view, are broadly similar to those of Varttala (2001), who analysed 60 research articles from economics, medicine, and technology, and found that the number of hedges in economics was about a third higher than in medicine and technology. In an analysis of 40 research articles from linguistics and medicine, it was found that while discipline did not greatly affect the proportion of hedging devices, it did influence the kinds of markers that were used (Vold, 2006). Thus while there may not be a consistent, and widely-agreed upon pattern of disciplinary variation when it comes to the use of hedging, it is clear that variation does exist.

The appropriate and effective use of hedging can present a number of challenges for L2 students. As Hyland and Milton (1997:184) state, 'Writer commitment can be expressed in an enormous variety of ways and these expressions can convey a wide range of meanings.' In the case of modals in particular, there are issues of polypragmatism, as these verbs can carry a

variety of meanings, which can refer to possibility, obligation, ability, or necessity, and have been described as vague and subjective (Hinkel, 2005, 2009; Hyland & Milton, 1997). Research has shown L2 students to favour certain modals over others, and to tend towards confident, rather than tentative predictions. Chinese students writing in English for example, were found to write with a more authoritative and direct tone, and to employ generally stronger modals than their L1 counterparts (Hu et al., 1982 – cited in Hyland & Milton, 1997). Flowerdew (2000) records a similar tendency amongst Hong Kong learners of English, finding that they stated conclusions as certainty, which resulted in an overly confident tone. This notion of L2 students having difficulties in finding the appropriate balance of confidence and tentativeness in their use of hedging has been raised in a number of studies. Bitchener & Basturkmen (2006), in a study of English L2 thesis writers, note that when it came to expressing the significance of their findings, and contextualising them in the wider research, students had a tendency to either over- or understate this, while other studies (Ädel & Erman, 2012; Chen & Baker, 2010; Gilquin et al., 2007; Hinkel, 2005; Hyland & Milton, 1997) observe L2 writers to underuse hedging devices or to use them less diversely than L1s, and to make more exaggerated claims. Conversely, in his study of Hong Kong students, Hyland (1994:140) comments on the finding that hedging was a frequent metadiscoursal device, demonstrating that the students showed 'a principal concern with expressing arguments explicitly and with due circumspection'. This may be due to the fact that the study focused on doctoral and masters students, the former of which at least one might presume to have significant previous experience in academic writing.

A number of studies have compared the use of hedging devices with that of intensifiers (such as, *greatly, totally, definitely, completely)*, which are common in speech and conversation.

Hinkel (2005) found that L2 writers demonstrated a generally lower frequency of hedging devices than did L1 writers, and conversely, employed intensifiers more frequently. Gilquin et al. (2007) draw attention to similar observations, noting that this can lead to register confusion. The reason for this tendency may relate to language exposure. As Hinkel (2005) notes, even L2 students who pursue language studies in English-speaking countries as a preparation for academic courses are exposed to far more informal, conversational input than they are academic register, which may explain their apparent familiarity with informal, spoken features, rather than those more commonly employed in academic discourse.

The tendency towards overstating claims and underusing hedging may also, in part, be a result of cultural background. Steinman (2003) draws attention to a number of differences in the academic writing norms of different languages and cultures, including varying perceptions of what constitutes evidence for a claim, how evidence should be presented, and how much of it is required. Lorenz (1998 – cited in Hinkel, 2005:34) comments on the varying uses of overstatement or hyperbole across cultures in academic writing. It may also be attributable in some measure to a form of overcompensation – if L2 writers feel that their language proficiency inhibits their ability to express their ideas, they may be more inclined to emphasise the importance of their message, even though this may negatively impact how appropriately they adhere to the norms of the academic discourse community (Hinkel, 2005). As Hyland & Milton (1997:183) observe, 'The ability to express doubt and certainty appropriately in English is a complex task for language learners'. It is nonetheless a crucial one, allowing writers to describe their findings and draw inferences from them without overstating their certainty or their importance within the context of the research field. This idea of contextualisation is also

central to another of the features of academic writing, and one with which, again, every student writer must become familiar if they are to write successfully – the practice of citation.

2.2.3.4 Citation

When writing for academic purposes, be it describing or documenting original research, or drawing together previous studies so as to examine a proposition or construct an argument, it is frequently necessary to refer to work that has been carried out by others. Indeed, Hyland (1999b) states that in academic articles in particular, referencing previous work is all but mandatory as a means of supporting current claims, as well as being a way to highlight one's own position. Citation plays a particularly important role in the literature reviews and discussion sections of research reports, as it allows the writer to illustrate how his or her study fits into the existing theory and knowledge base, why it is important, and how its findings can be interpreted and contextualised with reference to previous research. Citation helps writers to integrate the words and ideas of others into their own work, to provide context, and to be more persuasive when presenting their findings (Mansourizadeh & Ahmad, 2011). Citation is also a means by which writers can demonstrate their own subject knowledge and expertise, through critical evaluation of preceding studies. As Charles (2006:311) states:

> …it enables the writer to acknowledge or take issue with the contributions of other researchers and, in displaying knowledge of the field, to establish his/her own academic authority and credibility.

Hyland (2002b) notes that reporting structures come in four overall forms – direct quotes (either short, or more extensive in indented blocks), paraphrasing, summarising a single author, or generalising statements from a number of different sources. However, there are also important distinctions within this framework. Some of the early work examining the form and

variety of citations in academic discourse was carried out by Swales (1990). He produced a basic taxonomy, in which citations were either 'integral' or 'non-integral', and either 'reporting' or 'non-reporting'. For the purposes of clarity, it is worth briefly elaborating upon these descriptors. Integral citations are those where the name of the cited researcher appears as a syntactic element in the citing sentence itself, for example:

This is an integral citation, according to the taxonomy laid down by Swales (1990).

Or

Swales (1990) laid down a taxonomy for classifying citations.

Conversely, a non-integral citation would be one in which the cited author's name appears in parentheses or is referred to by a superscript number, as in:

Citations can be either integral or non-integral (Swales, 1990).

The second element of Swales' classification system refers to reporting or non-reporting citations. In this case, the former type would be those in which a reporting verb (such as *claim*, *state*, or *argue*) is used to introduce the citation. For example:

Swales (1990) proposed that citations could be reporting or non-reporting.

While the latter type employ alternative syntactic structures, not requiring the use of reporting verbs, as in:

Swales' (1990) classification of citations has largely stood the test of time.

This framework has subsequently been utilised in a number of studies examining the rhetorical choices involved in employing one type of citation or another. In comparing integral and non-integral citations, Hyland (1999b) proposes that integral citations appear to give greater

emphasis to the reported author, while non-integral citations place more of a focus on the findings being reported. Mansourizadeh & Ahmad (2011:157) comment on the 'objectivity and impersonality' that can be demonstrated by the use of non-integral citations, and it may also be that these are employed so as not to interrupt the flow of an argument (Hewings et al., 2010). It is not entirely clear however, whether writers always make these decisions consciously for rhetorical reasons, or simply to achieve variety in the way citations are presented. In the case of novice writers in particular, the manner in which citations appear is likely to have more to do with a lack of 'control over and awareness of how citation works in texts' than with predetermined rhetorical decision making (Jalilifar, 2012a:26).

Hyland (1999b, 2002b) takes the taxonomy of citations a step further, building on earlier work by Thomson & Yiyun (1991). He suggests that reporting verbs can be grouped in two ways according to function. Firstly, these verbs can describe different kinds of activity, so carrying out a 'process function'. This can include 'Research Acts' (*observe*, *calculate*, *analyse*); 'Cognition Acts' (*believe*, *assume*, *suspect*); and 'Discourse Acts' (*discuss*, *state*, *report*). Secondly, within this, the reporting verbs can also serve a more subtle 'evaluative function'. They are classified as 'factive', indicating that the writer represents the information reported as being true, or wishes to show the reported author in a positive light (verbs such as, *acknowledge*, *establish*, *verify*); 'counter-factive', suggesting that the writer regards the reported information as being false or wishes to represent the reported author negatively (verbs including, *ignore*, *overlook*, *exaggerate*); or 'non-factive', in which case the writer demonstrates a neutral stance towards the information or author being reported (verbs such as, *comment*, *state*, *point out*).

What then, does the literature tell us about how these various forms of citation are used by academic writers, and about how this might be affected, like so many other characteristics of academic writing, by disciplinary variation? As may be expected, discipline seems to affect the practice of citation in a number of ways (Samraj, 2013). Charles (2006) examines the use of reporting citations in theses written by L1 students in the disciplines of materials science and international relations. She finds differences between disciplines in terms of whether reporting clauses use human, non-human, or 'it' subjects, with the latter two being utilised with greater frequency in materials science, and in the matter of verb choice, where what she terms 'argue verbs' are most common in both disciplines, but occur with greater frequency in international relations. She also finds an overall predominance for integral citations, although this is in contrast to the results of other studies, possibly because Charles focuses only on reporting clauses, rather than all forms of citation. Non-integral citations were found to dominate among expert and novice chemical engineering writers at a Malaysian university (Mansourizadeh & Ahmad, 2011), and this has been noted previously as a tendency in the 'hard science' disciplines, with the exception of biology (Hyland, 1999b, 2002b; Okamura, 2008).

In examining a corpus of 80 research articles from leading journals in eight disciplines, Hyland (1999b, 2002b) makes a number of interesting observations, which are worth summarising here. Humanities and social sciences tended to make generally more liberal use of citations than did engineering or physics. In all of the disciplines, the most common form of citation was summary, followed by generalisation. Direct quotes were comparatively rare, and completely absent in fact from papers in science and engineering. There was also a general preference, across disciplines, for non-integral citations, with philosophy being the only exception. While the use of reporting verbs was higher in the humanities and social sciences than in hard

sciences, the variety of these verbs used was found to be more restricted in marketing, biology, and physics than in other disciplines. In terms of process function, Discourse Acts predominated in humanities and social sciences, whereas science and engineering disciplines favoured Research Act verbs. Non-factive verbs, those that indicate a neutral stance towards the reported information or author, were the most commonly occurring in all disciplines, with counter-factive verbs being relatively rare, and mainly confined to philosophy and sociology. Clearly then, not only does discipline play a role in determining the type and function of citations used, but there is also, potentially, a variety of rhetorical signals inherent in a writer's choice to cite in one way or another. While citation may, on the surface, appear to be a relatively straightforward element in academic writing, here too novice writers, and L2 writers in particular, may encounter difficulties. Moreover, citation, if not executed correctly, can lead to more serious problems for student writers than simply straying from register. Incorporating the work of others raises the potential, if the processes and conventions are not fully understood, for issues of plagiarism and academic misconduct.

In general, paraphrasing, summarising, and quoting, all of which are necessary elements in citing source material, can be especially challenging for L2 writers, particularly if they have not been trained in how to utilise source texts in the Western academic context (Hirvela and Du, 2013). In a study of Chinese postgraduate students in the UK, issues encountered included over-citation, repetitive formatting of citations and a limited range of reporting verbs, poor attribution of internet sources, and an over-reliance on directly copying source text (Davis, 2013). These difficulties are compounded by the various, and perhaps more subtle underlying characteristics of the different forms of citation – a complexity evidenced by the breadth of literature on the subject (Charles, 2006). For example, compared to writers of research articles,

MA thesis writers in Iran (writing in English) were found to employ a less diverse selection of factive verbs, and to utilise integral citations to a greater extent (Jalilifar, 2012a), a tendency also noted among novice L2 writers by Mansourizadeh & Ahmad (2011). While this may suggest that these writers were attempting to emphasise the reported author more than the reported findings, it may simply be that they were focusing on the grammar of the citations rather than any rhetorical purpose, and that the differences in practice really indicate 'unawareness of novices of the impact of citation types on readers and the interpretation assigned to the text' (Jalilifar, 2012a:36).

Other research has shown L2 students to use a generally smaller number of citations than L1 students when writing on the same subject, and to make much more frequent use of extended direct quotations (those over 40 words) (Borg, 2000). It is perhaps worth noting that while the same study, which examined the use of citation amongst L1 and L2 postgraduate students in the UK, also found that both groups of students experienced problems with the technical and rhetorical aspects of citations, L1 students did seem better able to express their stance towards the reported author or information than their L2 counterparts. Similarly, L2 undergraduate students have been found to adopt a neutral stance toward cited material, simply acknowledging it rather than taking a positive or negative position (Lee et al., 2018). This notion of employing more rhetorically complex citations types is potentially an important one. A greater reliance on citations that simply attribute information to an author rather than, for example, applying it to the writer's own work, or including an evaluative element, has been linked to lower scores in the context of MA theses (Petrić, 2007). This suggests that an understanding of the rhetorical aspects of citation use, above and beyond the idea of simply referring to previous research, is important knowledge for student writers to have. Underlining

this point, L2 writers of research articles in a variety of disciplines have been shown to utilise

subject-position integral citations more frequently than their L1 counterparts, again suggesting

that they may not be fully aware of the rhetorical purpose of this form of citation – to draw

attention to the author rather than the information reported (Okamura, 2008). Similar

tendencies among L2 student writers, to emphasise the reported researcher and to under-

employ evaluative citations have also been reported (Abasi et al., 2006; Jalilifar, 2012a; Samraj,

2013; Shi, 2012). In comparisons of student and published academic writing however, it is

important to remember that, as Petrić (2007) notes, these two groups of writers have different

audiences, different purposes for writing, and different levels of subject knowledge, all of

which may affect how citations are used.

Finally, it is necessary to address an issue related to citation use – that of paraphrasing and

plagiarism. If academic writers are not to employ direct quotations (which, as noted by Hyland

(1999b) are a relatively rare form of citation), they must find a way to express the reported

author's ideas, arguments, or findings in their own words. This requires an ability to effectively

summarise and paraphrase, as well as an understanding of when direct attribution to an author

is required. Here again, L2 writers in particular can encounter difficulties. L2 masters and PhD

students at three British universities were found to use source materials in ways that had the

potential to leave them open to accusations of plagiarism (Pecorari, 2006). With a more limited

vocabulary, L2 students may struggle to find their own words (Leki and Carson, 1997) or

syntactic structures with which to replace the original text. This may contribute to the fact that

L2 undergraduate writers have been found to struggle with generalisation and the synthesis

of multiple sources (Lee et al, 2018). They may also rely on methods such as simply substituting

synonyms, in the belief that this constitutes adequate paraphrasing (Shi, 2012). Strategies such

as this can place student writers at risk as they, albeit unintentionally, cross the 'unmarked borders of appropriate borrowing and lapse into unintentional plagiarism' (Shi, 2010:21-22). L2 students may lack an awareness of how to avoid these problems as a result of the complexity of skills such as paraphrasing (Shi, 2012).

Overall, the issues encountered demonstrate clearly that student writers, and in particular those for whom English is not an L1, will benefit, if they are to write successfully, from an understanding of the systems for acknowledgment, and the technical conventions, disciplinary variation, and rhetorical functions of citation, given its key role in academic discourse (Jordan, 2002). Indeed, the same need for a sound knowledge base can be seen in all of the features of academic writing examined in the preceding sections. For international students in the context of UK higher education institutions (HEIs), achieving an adequate IELTS score, and thus demonstrating a given level of general English proficiency, does not necessarily indicate that a student is fully equipped to deal with the challenges of academic writing. The fact that L1 students have also been shown to experience difficulties with academic discourse is evidence enough that the register is a complex one that can pose its own particular problems to the novice, and that these writers, especially when attempting to tackle academic discourse in an L2, may require pedagogic assistance specifically targeted at improving academic writing proficiency. EAP courses aim to provide this assistance.

2.3 English for Academic Purposes

2.3.1 What is EAP?

EAP is an eclectic discipline sometimes considered a subset of ESP (English for Specific Purposes), and incorporating 'classroom language, teaching methodology, teacher education, assessment of language, needs analysis, materials development and evaluation' among other

elements (Hamp-Lyons, 2011:89). EAP focuses on the 'teaching of English for use in academic contexts, to students for whom English is an additional language, and who are preparing to begin a course of academic studies...' (Thompson, 2006:2). It began to take form in the early 1960s at Birmingham University in the UK (Jordan, 2002) and its growth in the last few decades has been due in part to the increasingly prominent role of English in academia, including the fact that many disciplines have transitioned towards publishing in journals in English, rather than in those written in other languages (Jordan, 2002), as well as to a growing awareness that academic writing presents considerable challenges to international students in English-medium higher education (Wingate & Tribble, 2011). Indeed, writing forms a major element of the focus within EAP courses, and this has in turn led to increased interest in research examining academic discourse, research which itself then contributes back into the further development of EAP in terms of helping to clarify the features of discourse in different academic disciplines and genres (Coxhead & Byrd, 2007; Flowerdew, 2013).

EAP is now of major importance in countries such as the United Kingdom, Australia, and Canada, where large numbers of international students come to study. In the 2016/17 academic year for example, UK higher education enrolments by non-UK domicile students totalled 442,375 (HESA, 2018). Those students who are L2s often participate in pre-sessional EAP courses so as to be better prepared for the language requirements of their studies. Indeed, it is extremely important for these students to master the English they need in order to succeed in learning their subjects (Hyland & Hamp-Lyons, 2002). Academic writing is an vital element of this, and writing skill development is usually the subject of a strong focus in EAP courses (Bruce, 2005).

2.3.2 Research in EAP

The foci of research and development in EAP have been varied; there has been substantial discussion of the nature of EAP itself, particularly in terms of how best to deal with disciplinary variation, and whether EAP should be regarded as *pragmatic* - helping students to 'develop academic literacy skills to facilitate their effective participation in academic communities', or *critical* - questioning 'pedagogic and even political assumptions and values in the forum of our classrooms in order to provide learners with ways of considering their own academic socio-political status quo' (Hamp-Lyons, 2011:100). Attention has also been given to areas such as what might be thought of as more advanced applications of EAP, including writing in English for research publications, as well as to needs analysis and materials development (areas that I would argue are still of great importance, particularly in specific contexts such as pre-sessional university EAP courses). It is worth reviewing some of the research carried out in the field in attempts to address these issues.

2.3.2.1 The Nature of EAP

As EAP has developed, it has gone through a number of changes. Hyland and Tse (2007) argue for example, that in modern EAP, communication is given prominence over language, just as writing process is now considered to be as important as writing product. Hardy & Friginal (2016:119) highlight the fact that while some have argued for EAP instruction to be more general in nature, there has nonetheless been a movement towards specificity, 'emphasizing the need for students (and instructors) to become aware of disciplinary and genre practices outside of the language classroom'. Hamp-Lyons (2011) also notes the great progress that has been made in EAP research in the area of disciplinary variation. This provides a stark contrast to some of the earlier ideas expressed in the literature, such as those of Hutchinson and Waters

(1987 - cited in Hyland, 2002a), who contend that variation across academic disciplines in terms of grammar and discourse structures is insufficient to warrant subject-specific approaches.

As illustrated in the preceding sections, enormous variation in academic discourse has been highlighted not only between disciplines, but even between subdisciplines, despite the opinions of some lecturers, who believe academic writing to be largely homogenous (Harwood & Hadley, 2004). As an example of the impact this can have, students interviewed at two UK universities commented that they felt it was necessary to tailor their writing not only in accordance with genre, but also with subject and even with individual lecturers (Lea & Street, 1998). Similarly, Lea & Stierer (2004) note that different academic staff members give students different advice as to what represents quality academic writing. Harwood & Hadley (2004:356) encapsulate this situation by arguing that successfully acquiring academic literacy 'is a game with a bewildering set of rules, many of which are never made explicit to student writers'. The challenge for EAP, is how best to reflect this variation, and implement practices that, to continue Harwood & Hadley's analogy, might make the rules clearer.

Conceptualisations of EAP tend to fall somewhere on a spectrum ranging from *pragmatic* to *critical.* These two viewpoints see EAP very differently. Pragmatic EAP, which represents most standard EAP instruction, is instrumental, skills-based, and focuses on teaching students to be aware of the dominant conventions in academic discourse, so that they may utilise them in their own writing and thus cope with the demands placed upon them in higher education contexts (Harwood & Hadley 2004; Helmer, 2013). Conversely, critical EAP takes a wider view, questioning established hierarchies and power relations, as well as the notion that students should necessarily comply with accepted norms (Benesch, 2009; Helmer, 2013; Le Ha, 2009).

It is critical of the existing practices, and of pragmatic EAP's willingness to reinforce them, and instead promotes the idea that 'students have choices and should be free to adopt or subvert the dominant practices as they wish' (Harwood & Hadley, 2004:357).

While these may appear to be irreconcilable positions from which to approach EAP, efforts have been made to bridge the divide. Harwood & Hadley (2004) propose a 'Critical Pragmatic EAP' model. They argue that L2 student writers, particularly postgraduates, may move on to further research and attempt to publish their own work, and cite a study by Gosden (1992), which stated that editors of academic journals often receive such high numbers of submissions that they need scant reason to reject them, and that linguistic reasons are frequently enough. Given this, rejecting pragmatic EAP outright, and failing to teach the conventions of Western academic writing and its various disciplines may serve only to further disadvantage these student writers in the future. Nonetheless, they argue, the critical EAP standpoint has value in its contention that the varying and inconsistent expectations of lecturers when it comes to their students' academic writing should perhaps be questioned. They further propose that when considering conventions in academic discourse, it is reasonable to assume that there are some that can be safely ignored, and some that cannot, and suggest that a useful guide in establishing which conventions fall into which category might be provided by examining corpora, since this can provide useful insights into disciplinary practices.

Another approach, which in some ways parallels that of critical EAP in its emphasis on power relations and identity in student learning, is Academic Literacies (Lea & Street, 1998). The view here is that literacies are not a single set of transferable cognitive skills, but are rather social practices, the nature of which vary with context (Lea, 2004). Academic Literacies highlights 'the complexity of institutional and disciplinary requirements, which may be in conflict with

students' identities and previous experiences' (Wingate & Tribble, 2011:3). However, the applicability of this approach to L2 students in UK higher education has been questioned, as it is largely based on smaller-scale research, has not given a great deal of consideration to L2 students, and has not proposed very much in the way of principles and guidelines that can compare to the extensive research and disciplinary experience of EAP in this context (Wingate & Tribble, 2011). Lea herself (2004) acknowledges that the model has been challenged for its lack of a pedagogical design frame.

In considering my own views on the most appropriate approach to academic writing pedagogy, I would generally concur with Harwood & Hadley's *Critical Pragmatic EAP* model, as outlined above, and would certainly contend that corpus analysis has the potential to make clearer the importance (or lack thereof) of given conventions in academic writing across disciplines. As Nesi et al. (2004:446) argue:

> A corpus of student writing can serve as an awareness-raising tool for writing teachers in universities, schools, and colleges, and can thus accord with an academic literacies approach to writing instruction.

Therefore, importantly, this corpus-based methodology does not represent an outright rejection of any of the aforementioned approaches, but rather takes into consideration elements of them all. A balanced view of EAP approaches, combined with corpus analysis as a research tool may not only assist in informing which conventions might beneficially be taught to EAP students, but could also make clearer to both students and lecturers, which 'rules' need to be followed in which discipline, and which could be viewed as more negotiable – in this way acknowledging elements of all of the approaches outlined above.

It is also worth mentioning as this point, the not entirely unconnected notion of English as a Lingua Franca (ELF), and how this may or may not relate to EAP in the context of UK HEIs. ELF

refers to English when used as a medium of communication between parties who do not share a common first language, and given that L2 English speakers now outnumber those who have English as an L1, it has been the subject of much research. Some argue that it demands a re-evaluation of the way English is taught, and of the value placed on 'native speaker' standards. Bychkovska & Lee (2017:49) for example, argue that the criteria used to judge the writing of L2 university students in English-dominant higher education are still those designed for L1 students, and that 'variation from mainstream norms' of the language is still erroneously seen as a deficiency. However, it is important to account for context. While in terms of everyday communication between international L2 English students, higher education in the UK may be an ELF context, there are many regards in which it is not one (communicating with L1 student colleagues or staff, and submitting written work to L1 lecturers for example). As Björkman (2011) notes, EAP for international students studying in English-speaking countries (Chinese students in the UK, for example) aims primarily to improve their receptive and productive skills in preparation for their studies, and thus ELF may not find its greatest relevance in this context.

More widely, even where EAP is taught in EFL environments, there is still an apparent preference among teachers and particularly students for the 'native speaker' model, as demonstrated by Griffiths & Soruç's (2019) review of multiple studies from countries such as Greece, Argentina, Japan, Finland, and Germany, and by other research from Turkey and China (Kaypak & Ortaçtepe, 2014; Zhang and Du, 2018). While some researchers argue that this indicates a lack of awareness of ELF, and suggest that learners' re-evaluate their beliefs about learning English (De Bartolo, 2018), others contend that even English for general purposes should be taught as a 'fixed set of codified norms', since language cannot function as an effective means of communication without these (Swan, 2017). The centre ground here is

perhaps that occupied by Maley (2009), who points out a number of problems with the 'strong'

ELF model, and suggests that learners awareness of ELF be raised, while still actually teaching

what amounts to a standard variety of the language. Given the particular demands of academic

writing, the many potential conventions of this genre, and the intended audience of the

ultimate outputs that result from it, I would tend towards the opinion that teaching a standard

variety of English may currently be the approach best suited to EAP writing instruction.

2.3.2.2 EAP Needs Analysis

Needs analysis is, in principle, an important element of all language education courses, as

success depends on meeting the needs of the participants; this applies equally to EAP and the

teaching of academic writing (Berman & Cheng, 2010; Hu, 2007). As Carkin (2005:87) observes,

'needs assessment of the diverse learners in EAP underlies syllabus design, materials

development…and, ultimately, evaluation of students and course or program success'. For this

reason, a great deal of attention has been given to needs analysis in the EAP context, in an

effort to determine what it is that learners most need to learn, and by association, how

materials, teachers, and courses can best meet these needs. Nonetheless, in terms of needs

analysis the university EAP context is not a simple one, and the many and varied demands

placed on students, especially L2 postgraduate students, mean that answering the question of

what these learners need can be a complex proposition.

 Studies have produced a variety of results when it comes to EAP needs analysis. In their

longitudinal study of undergraduates at an English-medium university in Hong Kong, Evans &

Morrison (2010) state that students found academic writing to be their main difficulty, with

particular reference to the use of specialised vocabulary, uncertainty over the academic

requirements of their professors, referencing, differing genres, extended writing, and

appropriate style and register. However, they also contend that the central concern of EAP courses should be to assist students in understanding the discourse practices of their disciplines. This parallels the idea that it is important for EAP materials to relate to specialised knowledge, particularly when it comes to subject-specific vocabulary, a point that has been made in a number of previous studies (Abdolrezapour & Tavakoli, 2013; Berman & Cheng, 2010; Chan, 2001; Evans and Green, 2007; Evans & Morrison, 2010; Leki & Carson, 1994). If students are to be prepared for the specific demands of their disciplines, non-specialised vocabulary instruction may prove inadequate. Indeed EAP students have been found to show a preference for activities that are oriented towards specific disciplines over activities such as general study skills (Liyanage & Birch, 2001).

Another perceived need, particularly when students themselves are surveyed, is grammar. Hinkel (2003:299) argues that teachers of students who will go on to academic study should consider ways in which to furnish their learners with 'more sophisticated syntactic constructions', while Hu (2007) notes genre-specific grammar as one of the elements of knowledge considered in the design of an EAP course in Singapore. Similarly, grammar represents the predominant language component of EAP in Malaysia (Shing & Sim, 2011), and indeed, grammar has been shown to be a concern, and a perceived need, by both academic programme leaders and students (Evans & Green, 2007). Nonetheless, it is also necessary to bear in mind that, as Evans & Morrison (2010) point out, university lecturers may be more interested in the content of students' academic writing than they are in grammatical accuracy. They may also pay relatively less attention to grammar mistakes and issues at sentence level (Leki & Carson, 1994), a notion which Hinkel (2009) appears to echo when he proposes that examining the syntax and lexis of L2 writing cannot provide a comprehensive indication of its

overall quality. So while there is certainly value in focusing on grammar and study skills, other elements of academic writing must also be considered.

Tanaka & Gilliland (2017), for example, argue for critical thinking instruction in EAP, on the grounds that this would promote the consideration of multiple viewpoints and allow students to produce well-reasoned arguments. This would indeed address one aspect of academic writing content, however another valuable approach may be to consider some of the other linguistic features of academic writing. Specialised, discipline-specific vocabulary has already been highlighted above, but research has also pointed to the potential benefits of instruction in other areas. Petrić (2007) notes that rhetorically varied citation use is associated with higher scores in writing, and suggests that greater attention be paid to this in EAP classes. A lack of set rules where the use of citation is concerned means that instruction may be of benefit to students (Shi, 2010), who should understand how expert writers employ citation practices. Informed teaching can help to make this more explicit, and ideally should serve not only to emphasise to students that plagiarism must be avoided, but also teach them how to cite appropriately (Pickard, 1995; Hu, 2007). The potential value of this is demonstrated by Wette (2010), who observed an increase in the proficiency with which L2 student writers employed citation practices after a relatively short period of instruction. Another area that has been highlighted as one to which attention can usefully be paid in the university EAP context is hedging, particularly given that more frequent use of these devices has been associated with both higher L2 student proficiency and higher essay scores (Ferris, 1994; Friginal et al., 2014). An understanding of hedging can practically benefit textbook design (Hyland, 1996), and writing courses can assist students by emphasising not only the role that hedges play in academic discourse, but also those items such as intensifiers, which students may benefit from

avoiding (Hinkel, 2005). Numerous studies have also espoused the value of EAP pedagogy that includes a focus on lexical bundles. Attention can be drawn to those bundles that commonly occur in expert writing, so as to assist learner writers (Chen & Baker, 2010). Moreover, as learners may not always find it easy to acquire an understanding of the precise discourse functions of such bundles, even if they encounter them, instruction in their use may also be of benefit, particularly in activities based on communicative writing (Biber & Barbieri, 2007; Pang, 2010). Finally, it is important to bear in mind, as Jarvis et al. (2003:399) argue, that 'the quality of a written text may depend less on the use of individual linguistic features than on how these features are used in tandem' and as such, EAP students in the university context may need balanced course content that gives attention to a variety of features and attempts to view them holistically to some extent.

 Despite the fact that research has already shed a great deal of light on the needs of EAP writing students, it must be remembered that need analysis itself requires regular re-evaluation (Belcher, 2006). It should be viewed as a process that is ongoing, rather than one carried out only when a course is first being designed (Abdullah, 2005). Ongoing needs analysis can help to determine whether the needs of learners are being met, and can highlight potential areas for change (Liyanage & Birch, 2001). If a course aims to bridge the gap between the current situation and the desired goals, it is important to know what that situation is, and what those goals are (Graves & Xu, 2000). In the context of EAP courses, there are a variety of ways in which this needs analysis might be conducted.

 With its central role in EAP, needs analysis research has often focused on the needs and preferences of students, and the expectations and demands of lecturers, often through faculty surveys (Evans & Green, 2007; Leki & Carson, 1994), and as shown, studies have highlighted

various priorities in terms of the needs of EAP learners, not always reaching similar conclusions. At a general level, it has been argued, EAP courses aimed at those who will go on to university study should be academically focused, should nurture an awareness amongst students of the academic language used to write about texts and ideas, and should consider the target students' language needs (Jordan, 1997; Liyanage & Birch, 2001; Shing & Sim, 2011). The issue arises when one attempts to determine precisely what those language needs are, and how to best create the necessary awareness of academic language. This must take into account disciplinary variation; the demands placed on students by differing genres, and indeed lecturers; and the fact that pre-sessional EAP courses often see students who will go on to very different disciplines studying academic writing in the same class. Needs are determined by a combination of factors, including among others, level of study, discipline, and specific subject area (Berman & Cheng, 2010).

Leki & Carson (1994) suggest various means by which the needs of EAP students may be analysed. These include surveying the kinds of writing assignments required of them, the rhetorical skills students will need in order to complete those assignments, the attitudes of academic staff towards the writing produced by students, and the perceptions of the students themselves. The analysis process can also be viewed in terms of target situation (what learners will need in order to successfully study within their discipline), present situation (learners' strengths, weaknesses and language proficiency), and learning situation (learners' motivations, learning strategies/styles, and expectations) (Abdullah, 2005; Dudley-Evans & St. John, 1998). Yet another alternative is to examine the perceptions of EAP teachers and students in terms of the efficacy of the methodologies used in EAP classes themselves (Abdolrezapour & Tavakoli,

2013). Overall, it may be that utilising a combination of some of these methods is the optimal approach to adopt.

While a great deal of the work carried out on academic writing has been quantitative in nature, there have also been qualitative studies examining the writing itself, as well as the ways in which EAP pedagogy might be informed. Chinese students' understanding of the function of paraphrasing was examined through interviews and think aloud protocols (Hirvela & Du, 2013), while interviews were conducted in order to investigate Canadian L2 and L1 doctoral students' citation use in grant proposals (Fazel & Shi, 2015). The views towards paraphrasing, summarising, and plagiarism among L2 students and their instructors at a North American university were also studied through the use of interviews (Shi, 2012). EAP students' needs and attitudes when it comes to their instruction have been examined too. A questionnaire survey was carried out to establish Hong Kong university students' perceptions of their language needs (Chan, 2001), while interviews were used to investigate the EAP and university writing experiences of L2 students in the US (Leki and Carson, 1997), and the questionnaire format was once again employed for a study into the perceived language difficulties of L2 students in a Canadian university (Berman and Cheng, 2010).

A number of studies have combined quantitative and qualitative methods, something that is advantageous in its potential to provide a degree of data triangulation. Li and Schmitt's (2009) study centred on an individual Chinese MA student in the UK, and employed both interviews and text analysis in order to understand her use and acquisition of lexical phrases. Davis (2013) studied citation use by three Chinese postgraduate students in the UK, again combining the analysis of texts and the use of semi-structured interviews, and a similar research design was utilised by Petrić (2012) in order to examine L2 students' use of direct quotation in a corpus of

16 high and low-rated MA theses. Studies employing this mixed methods approach may benefit from the insights available when objective, quantitative data and subjective, qualitative data are examined in combination. A study such as this in the university EAP context has the potential to offer valuable insights into academic writing, and allow pedagogical implications to be clearly highlighted.

For EAP writing instruction to be effective, academic writing instructors should be aware of the demands their students will face in their other courses when it comes to the linguistic and rhetorical features of academic discourse (Hardy & Friginal, 2016:119), given that EAP writing classes aim to furnish students with the knowledge they need in order to write more proficiently in their academic subjects, rather than within the confines of the EAP writing classes themselves (Leki & Carson, 1997). By way of illustration, Andrade (2006) refers to the case of international students in the UK enrolled on a master's programme in business administration, who, while they had been given additional materials to assist with their English, reported that they felt these were unhelpful as they dealt with general topics rather than being focused on content specific to the discipline they were studying. In considering the issues of disciplinary variation and teaching academic writing, Hyland (2002a:389) goes so far as to propose that ignoring specificity undermines pedagogic effectiveness, and argues that 'The discourses of the academy do not form an undifferentiated, unitary mass but a variety of subject-specific literacies. Disciplines have different views of knowledge, different research practices, and different ways of seeing the world'. International L2 students coming to study at universities in the UK must necessarily develop various skills in academic writing (Read, 1990), and the challenges of this are potentially many. This may be especially true in the case of those students who are unable to demonstrate the required language proficiency for

acceptance to the university and must therefore undertake an pre-sessional EAP course before they commence their studies proper.

2.3.2.3 EAP Pre-sessional Courses

With the increasing numbers of international students in recent decades, pre-sessional English courses, which aim to assist L2 students in meeting the linguistic demands of their programmes of study before they commence, have also witnessed an expansion (Moore & Morton, 2005). A student may derive little benefit from their university studies if they face a serious struggle with the language used in the institution, and as a result, UK universities usually require some evidence of English proficiency (Green, 2007) in the form of a score on the academic IELTS test. If an applicant fails to attain the required score, they may be accepted on condition that they complete a pre-sessional English course prior to the start of their academic program. In the UK, these courses tend to have durations of between 5 and 20 weeks, depending on the institution, admissions procedures, and IELTS scores attained by the prospective students, and are generally run over the summer period, prior to the commencement of the autumn term in September. Pre-sessional English courses are designed to assist L2 students in getting to a point where, linguistically speaking, they are ready to start their courses, although they may also serve to familiarise students with the environment and facilities of their chosen institution.

It is clearly important that these courses be as effective as possible, particularly in light of the fact that for many international students, studying in the UK is a considerable financial commitment, with tuition fees for non-EU students being substantially higher than those paid by UK or EU students. For these individuals, optimising their outcomes while engaged in study in the UK is likely to be a major consideration. Likewise, higher education institutions themselves seek to be successful in terms of student outcomes and completion of courses.

There is therefore value in any tool that can be utilised to help ensure that pre-sessional materials and course content are as beneficial as possible in accurately reflecting the features of academic discourse and the variation across disciplines, particularly if this can assist in informing which features of academic writing can most usefully be included in these courses. One approach that has already proved to be of great utility in informing EAP in general is corpus linguistics.

There do already exist EAP textbooks and teaching materials that are corpus informed, such as the Cambridge Academic English series, which is compiled with reference to the Cambridge English Corpus (Cambridge English, 2018). However, the extent to which even these materials address disciplinary variation in features such as vocabulary, hedging, and citation is limited, as by their nature they are designed to be suitable for mixed-discipline groups and to teach 'language and skills that will be of use to students working in all subjects' (Hewings, 2012:5). In any case, it is equally true that in specific pre-sessional contexts, teaching materials may be produced in-house, and thus may not benefit from the insights provided by the analysis of academic corpora.

2.4 Corpus Linguistics

2.4.1 Corpus Linguistics-An Introduction

Kennedy (1998:1) defines a corpus as 'a body of written text or transcribed speech which can serve as a basis for linguistic analysis or description'. Corpus linguistics (CL) involves 'the collection and analysis of large amounts of naturally occurring spoken or written data in electronic format' (Gilquin et al., 2007:320), and as Thompson (2006) notes, a corpus should also act as a sample of kinds or varieties of discourse, as well as being the basis for the analysis of linguistic features. In corpus linguistics, this written or spoken data is collected and analysed

so as to study grammatical and lexical features, in order to ascertain, for example, how commonly they are used, how they co-occur, or which of them are typical of a given genre. CL constitutes a powerful analysis method (Gilquin et al., 2007; Nesi et al., 2004).

Early CL research tended to use large-scale, generalised corpora to focus on grammatical and lexical items, and it was subsequently employed in areas such as historical linguistics, teacher education, and translation (Flowerdew, 2002). Over time, its utility in language teaching also became apparent. As McEnery & Wilson (2001) point out, corpora can help to raise awareness of the grammatical features and vocabulary that are used in authentic discourse; this means that corpora are potentially of value to a variety of users, including grammarians and lexicographers, syllabus and materials writers, researchers, language testers, teachers, and learners (Thompson, 2006). CL is now well established as 'a powerful methodology-technology' (Lee & Swales, 2006:57), and one that can be employed in various roles, examining areas such as lexis, grammar, semantics, pragmatics, sociolinguistics, and stylistics, as well as contributing to language teaching (McEnery & Wilson, 2001).

2.4.2 Corpus Linguistics and Academic Writing

Various studies note that data derived from the analysis of corpora has contributed to EAP descriptively, as well as in terms of academic writing pedagogy (Chang & Kuo, 2011; Hardy & Römer, 2013). It has allowed researchers to determine the features of academic discourse (Krishnamurthy & Kosem, 2007; Lee & Swales, 2006), and has brought into doubt the notion of a single academic register, vocabulary, or style across all disciplines (Tribble & Wingate, 2013). As Gilquin et al. (2007:321) note, 'The research paradigm of corpus linguistics is thus ideally suited for studying the linguistic features of academic discourse as it can highlight which words, phrases or structures are most typical of the genre and how they are generally used'.

A wide variety of studies has been conducted using the analysis of corpora to explore the nature and features of academic writing, for both pedagogical and descriptive purposes, and this has also involved the creation of a number of corpora for the purpose. Sometimes corpora are built for a specific study, while in other instances they may be created with the goal of producing a resource of use to a variety of researchers for a variety of purposes. The purpose for which a corpus is created necessarily influences what it includes and how it is designed.

The corpora compared in the study of academic English and EAP therefore take various forms. Sometimes L1 student writing is compared to L2 student writing, while in other studies 'expert' or published academic writing is used as a point of comparison. The number of academic disciplines included may also differ, as well as other factors such as the level of the students (undergraduate, master's level, or PhD for example), the proficiency of the writing, and the genres analysed (research reports, argumentative essays, thesis introductions, examination scripts, textbook chapters, and so on). Obviously the analysis of academic corpora is not a 'one size fits all' situation – how corpora are built or which elements of them are utilised depends very much on what one wishes to get from an analysis, and what questions a given piece of research wishes to ask. A review of some of the previous work undertaken using corpora to gain a clearer understanding of the features of academic discourse, and how this can be applied to EAP, will serve to illustrate these various considerations, and how corpus design and use may change in response to the differing requirements of researchers. It will also demonstrate the important role CL has to play in understanding, and by extension, teaching academic writing, as well as helping to further situate the present study in relation to other research utilising the CL approach in the EAP context.

2.4.2.1 Source Writers for Corpora

Much of the work done with corpora in the area of EAP has involved comparing learner corpora with those made up of parallel L1 writing or of writing considered 'expert', such as published research articles, textbooks and so forth (Flowerdew, 2002). Exactly whose writing is used in a corpus study however depends on the specifics of the research, and sometimes on individual researchers' opinions, as there is not always universal agreement in terms of what constitutes appropriate sources when analysing corpora, even when studies share similar goals. This is particularly true when it comes to sourcing writing suitable for comparison to that of L2 student writers.

In some cases, the decision about which writers to source from is a relatively straightforward one, and a number of studies have focused solely on published academic writing. For example, in her study into citations and quotations, Pickard (1995) aimed to examine how these features are used by accomplished writers. She therefore chose to look at research articles, on the grounds that the authors of these articles can be considered 'expert' in that they have succeeded in publication and have thus been accepted into an academic discourse community. Similarly, a corpus consisting of research articles written in L1 English, L2 English, and L1 Spanish was analysed in order to investigate the use of formulaicity by academic writers across language variables (Pérez-Llantada, 2014). In this case the selection of these source authors was designed to ensure familiarity with conventions such as academic register and style. This model is not unlike that employed by Jalilifar (2012b), in a study examining research article introductions in international and Iranian journals. A corpus upon which a number of studies has been based is that compiled by Hyland (1999b, 2002b). This was used to explore conventions in how citation use varies from discipline to discipline, and as such necessarily

consisted of 'expert' writing, in this case 80 research articles from a variety of disciplines. In another study by Hyland (1994), the corpus comprised texts from 22 textbooks designed for L2 EAP students. Here the intention was to evaluate how these books addressed the issue of hedging, and how this compared to conventions established in the literature. For further examples of studies employing a model of using only 'expert' writing, see Hewings et al. (2010), Martínez (2005), Okamura (2008), Shahriari (2017), and Vold (2006).

Other studies have focused primarily on examining student writing, in some cases looking solely at either L2 or L1 students. Pecorari (2006) used a corpus containing sections from drafts of dissertations and theses written by L2 PhD and master's students to investigate the extent to which novice writers adhere to the expectations of their disciplines when it comes to citation. In this case, disciplinary expectations were determined by reference to existing research, and through interviews with students and their supervisors, rather than by direct comparison with an 'expert' corpus. In another example, a corpus consisting only of L2 student writing was compiled so as to examine differences in metadiscourse between two groups of L1 Mandarin speakers writing in English – those in an English-medium environment in China, and those studying in the UK (Li & Wharton, 2012). Variation in the use of direct quotations between high and low-rated MA theses was examined by Petrić (2012), with a view to providing recommendations for academic writing pedagogy. In this instance too, the source writers for the corpus were solely L2 student writers. Zhao and Llosa's (2008) study serves as an example of corpus research in which only L1 student writing was analysed. The corpus consisted of 42 English Language Arts examination papers, and was used to investigate the role of voice in L1 student academic writing, so as to provide data that might facilitate future comparison between L1 and L2 writing.

The examples given in the preceding paragraphs make it clear that corpora comprised solely of writing from 'experts', or from either L1 or L2 student writers can be of great value in investigating various aspects of academic writing, be this, for example, variation across languages, levels, disciplines, or academic rating success. However, when comparisons are made between two or more of these writing sources, corpora can begin to provide many more insights that can be applied to EAP pedagogy, particularly by highlighting differences in the usage of academic discourse features between L2 students and either their L1 counterparts or 'experts' in their academic disciplines. Sometimes research aims and study design make the choice of source authors a straightforward one, but this is not always an unproblematic issue. This is true particularly in cases where the aim is to compare L2 student writing to 'model' academic writing. The question becomes, who provides the best model, 'expert' writers, or L1 students? Research in the past has approached this in different ways, and with differing justifications.

A number of studies have compared L2 student writing to that of L1 students. Hyland & Milton (1997:184) investigated the use of hedging and how doubt and certainty are expressed by student writers. The study involved a corpus of exam scripts from Hong Kong Chinese and British students, writing in English at A-level, and aimed to 'determine the typical forms and meanings used by the two groups to present claims in academic English prose.' In another study, citation practices were examined through a corpus of essays by five L1 and eleven L2 students at a British university, with a view to comparing the difficulties experienced by these two groups of students (Borg, 2000). Friginal et al. (2014) conducted a study into the co-occurrence of linguistic features and how this influences the profiles of successful under and postgraduate L1 and L2 academic writing. The corpus used was sourced from 353 L2 and 150

L1 students in the USA. Rather than considering the L1 writing to be a benchmark of quality against which to evaluate issues in L2 writing, the study aimed to investigate how highly rated essays by the two groups might differ in terms of style and characteristics.

While these studies do not assume L1 student writing to be 'model' writing for comparison, other research has viewed this differently. Granger & Tyson (1996), in a study of connector usage in L1 and L2 English essays, hypothesise 'overuse' of connectors by L2 writers. Since the comparison corpus is L1 essays, this would seem to imply that these must represent 'appropriate' usage, and that therefore, in this sense at least, these students represent 'model' writers. Similarly, Hinkel (2001) refers to excessive use of demonstrative pronouns in L2 students' essays when compared to those of L1 students, again suggesting that the latter can automatically be regarded as demonstrating usage that is entirely appropriate. This notion of using L1 student, or 'novice' writing as a model against which to compare the writing of L2 students has however been questioned. Gilquin et al (2007) argue that, with particular reference to informing pedagogy, the conventions observed in novice L1 writing may not necessarily be desirable, or represent models worthy of imitation. Similar points are made by Bolton et al. (2002) and by Ädel (2006:206-207):

> …it can be argued that in order to evaluate foreign learner writing by students justly, we need to use native-speaker writing that is also produced by students for comparison. On the other hand, it can also be argued that professional writing represents the norm that advanced foreign learner writers try to reach and their teachers try to promote. In this respect, a useful corpus for comparison is one which offers a collection of …'expert performances'

A number of studies have adopted this approach, but here too, comparisons may not be entirely unproblematic. The introductions of 65 L2 MA theses and 65 research articles in applied linguistics were used by Jalilifar (2012a) to investigate citation use. The study contends

that knowledge of how expert and novice citation use differs can help to guide novice writers of English by making them more aware of how this feature of academic writing is employed by those who are familiar with discourse community conventions. Interestingly however, Jalilifar also acknowledges that some of the observed differences may be attributable to the fact that research articles and MA theses (and by extension other student writing) have quite distinct communicative purposes. This point is also central in Hyland's (2008b) study of lexical bundles in research articles and L2s' MA dissertations and doctoral theses. Rather than assuming that any differences in the use of lexical bundles reflect deficiencies on the part of the L2 students, Hyland notes that these different genres are aimed at different readers and have different purposes, and that this may well shape how language is used within them. As Gardner & Nesi (2012:28-29) observe:

> …research genres aim to persuade the reader of the validity of new findings…assignments generally aim to demonstrate the acquisition of required skills and accepted knowledge.

Therefore, assuming differences to indicate an automatic dichotomy of the 'rightness' of published academic writing and the 'wrongness' of L2 students' writing is perhaps ill-advised. The results of any comparisons between novice and expert academic writing, and any potential pedagogical applications emerging from them must therefore be carefully considered.

Other objections to comparing expert published writing with that of L2 students have also been made. Hyland & Milton (1997) contend quite specifically that it is unrealistic to measure the efforts of L2 students against those of 'expert writers', particularly as published work has often been repeatedly reviewed and revised. Thompson (2006:5) argues that corpora of student writing can be more useful than those of expert texts, since the academic writing of students can be best informed by data derived from texts that are 'analogous to those that

they are expected to produce'. This point is echoed by Harwood (2005) in an evaluation of the corpus research used to inform EAP textbooks. He contends that for EAP materials to be based solely on how journal articles are written is oversimplistic, as the genres of student and published academic writing are different, and the practices of expert writers may therefore be of limited relevance to students. As similar contention is made by Hardy & Friginal (2016:120), who argue that while research articles are a product of the 'inner circles of their respective discipline's centre', students can be viewed as being somewhere on the periphery, such that for them, learning to write in the same way as experts may be unrealistic. Harwood (p.157) concludes that corpus analysts should employ a 'battery of generically diverse student and expert corpora' if EAP materials are to be evaluated. I would add that these corpora should also be multidisciplinary in nature. How disciplines are represented in academic corpus research is the next area to which attention will be turned.

2.4.2.2 Disciplines in Corpora

The extent to which varied disciplines are represented in academic writing corpora is also very much dependent on why a given corpus is compiled, and for what it is intended to be used. Even in the cases of corpora compiled for general research purposes rather than to facilitate a specific study, the representation of different academic disciplines varies. Some corpora seek to present a picture of general English, rather than that used in specific academic disciplines. An example of this is the International Corpus of Learner English (ICLE, Granger et al., 2002) This 3.5 million word corpus consists of 6000 mostly argumentative essays from 16 L1s, written on a wide variety of topics, and was tagged to allow for the analysis of learner errors. One of the first uses of this corpus was to investigate word frequencies so as to establish patterns of over or underuse. In order to achieve this, Granger & Rayson (1998) compared a 280,000 word

section of the ICLE, made up of formal argumentative essays, with a 230,000 word section of the Louvain Corpus of Native English Essays (LOCNESS). The ICLE was also utilised by Gilquin & Paquot (2007), who selected from the corpus as a whole, only untimed argumentative essays written without the use of reference tools. This sub-corpus amounted to around 1.5 million words from 14 L1s, and was compared to the spoken and written academic components of the British National Corpus (BNC), with the aim of assessing the degree to which learners' written English employs features more usually indicative of speech. Gilquin & Granger (2011) made further use of elements of the ICLE to investigate the differences between ESL and EFL learners with regard to their usage of 'into'. Here the learner corpus was compared to the LOCNESS and two corpora of English newspaper editorials.

The Hong Kong University of Science and Technology Learner Corpus (HKUST), was developed along similar principles, with essays from EFL courses, written by students from a variety of disciplines. A 500,000 word interlanguage extract of this was compared to high-scoring UK general studies exam scripts in order to examine differences in the over and underuse of four-word formulaic sequences (Milton, 1998). These same corpora were also utilised in Hyland & Milton's (1997) study into how doubt and certainty are expressed in academic writing.

Clearly then, the uses to which general corpora such as those detailed above can be applied are plentiful. However, if one wishes to examine academic writing more specifically, for example investigating areas such as subject-specific vocabulary, citation practices in given academic discourse communities, how lexical bundle use varies between science and humanities subjects, or indeed any other aspect of disciplinary variation, corpora that include writing from specific disciplines are required. One of the first of these was the JDEST Corpus, compiled in the mid 1980s at Shanghai's Jiao Tong University (Thompson, 2006). While

technical in nature overall, this corpus contained work from 10 disciplines including physics, chemical engineering and computer science. Another corpus created for general use but incorporating specific disciplines was the British Academic Written English (BAWE) Corpus. This was developed by Reading, Oxford Brookes, and Warwick universities, in order to 'enable the identification and description of student writing genres across disciplines and at different stages of academic development' (Alsop and Nesi, 2009:72). As such, the BAWE Corpus contains university assignments from 30 different disciplines. Researchers at the University of Michigan produced the Michigan Corpus of Upper-level Student Papers (MICUSP, Römer & O'Donnell 2011:159) to allow 'corpus researchers, EAP teachers and language testers to investigate the written discourse of proficient, advanced-level native- and non-native speaker student writers'. The corpus incorporated a wide range of university papers, (research reports, argumentative essays, proposals, creative writing and so on) from 16 different disciplines across Social Sciences, Physical Sciences, Biological and Health Sciences, and Humanities and Arts. The MICUSP has since been used in a variety of studies, focusing on issues including cross-disciplinary use of the pronoun 'this', the lexico-grammatical features indicative of highly rated student writing, disciplinary variation in citation practices, and German learners' use of the progressive in English writing (Ädel & Garretson, 2006; Hardy & Römer, 2013; Römer & Wulff, 2010; Wulff & Römer, 2009).

Many corpora are compiled for specific studies rather than as general resources, and in these cases, disciplinary representation is closely tied not only to the purpose of creating the corpus, but also in some cases, to practical considerations. Pecorari's (2006) study into citation use and disciplinary expectation in L2s' MA and doctoral theses included four disciplines – biology, linguistics, education, and civil engineering. While this does provide a range across which to

examine differences, the study makes clear that the inclusion of these disciplines resulted from which students and supervisors agreed to participate in the research project. Peacock (2002) offers differing justification for his inclusion of seven disciplines - Physics and Material Science, Environmental Science, Language and Linguistics, Biology, Law, Business (Marketing and Management), and Public and Social Administration – in a study of disciplinary and L1/L2 variation in research article discussion sections. In this case, inclusion was based in part on representing academic range, in part on selecting subject areas in which there are considerable numbers of L2 students (so as to provide useful pedagogical recommendations), and in part to the fact that these disciplines had been under-represented in previous similar research. Although not specified, providing suitable range was presumably also Moore & Morton's (2005) primary consideration in selecting the number of disciplines to include in their corpus study comparing university and IELTS writing. The 26 subjects, ranging through areas such as accounting, anthropology, physics, history, and tourism, would provide a reasonable representation of the various courses L2 students might embark upon, and would thus be a good basis against which to examine the differences and similarities between the writing done in university, and the 'academic' IELTS tasks.

Obviously, not every corpus study focuses on or requires wide disciplinary range (see Jalilifar, 2012a; Mansourizadeh and Ahmad, 2011; Okamura, 2008; and Samraj, 2013 as examples), but in those studies that do seek to incorporate a broad variety of academic disciplines into their corpora, this is accomplished using a variety of models and frameworks. Hyland (1996b:344-345, 2002b) used a corpus of 80 research articles from 10 journals across eight disciplines ('molecular biology, magnetic physics, marketing, applied linguistics, philosophy, sociology, mechanical engineering, and electronic engineering') to examine citation practices, while using

56 research articles from eight similar though not identical disciplines to investigate writer stance (Hyland, 1999a). Disciplinary range was represented in a different way by Vázquez and Giner (2008). Their study into hedging in research articles incorporated papers from journals at three points along the disciplinary continuum – soft sciences being represented by marketing, and hard sciences by mechanical engineering, with biology occupying the middle ground between the two. Paquot (2007:130), in a study aimed at identifying a productively focused academic word list, necessarily required a wide range of disciplines, and thus utilised a corpus including 'arts, belief and religion, applied science, science and social science'.

What is clear therefore is that as with other aspects of compiling and using corpora, the representation of academic disciplines is a matter that depends on a number of different considerations, some related to more practical issues such as the nature of conducting research within academic institutions, and others stemming from the requirements of the research itself. For a study aiming to examine the potential to inform pre-sessional EAP content, the ideal corpus design, given the wide range of university courses on which pre-sessional students may go on to study, would be one with a wide disciplinary range. With that said, it is also inevitable that this must be tempered to some degree by the real-world practicalities of accessing source material.

2.4.2.3 Genres in Corpora

Another decision that must be made when compiling corpora, relates to the text types or genres that will be included. Academic writing can exist in a wide variety of forms, such as research articles, textbooks, dissertations, theses, book reviews, essays, research reports, proposals, and so on. These different genres may also be further subdivided; previous research for example has included studies examining only the introduction, abstract, or discussion

sections of research articles or theses, or individual chapters from textbooks (Ozturk, 2007; Paquot, 2007; Ren & Li, 2011). Once again, the choice of genres included in any corpus analysis is determined by what that analysis aims to achieve.

The choice of using published or student writing has, to a large extent, been addressed in section 2.3.2 (i), so here the issue of genres within those broad delineations will be focused upon. One of the primary concerns is whether to make use of pieces of writing in their whole form, or simply to analyse extracts or sections of a given sample. There are studies applying both of these approaches to text integrity, and again, the decision will usually be based on the aims of the research in question.

In some cases, research sets out with the stated aim of examining only specific sections of research articles, theses, or whichever text type has been chosen. Genre analysis studies will often target specific sections of texts, such as in the case of Brett (1994). This study examined the move structure of the results sections of sociology research articles, so as to provide a description with pedagogical applications. Similarly, Kwan (2006) conducted genre analysis of a corpus of literature reviews from doctoral theses so as to compare this to an established model for introductions, while Yakhontova (2006) investigated cultural variation in conference abstracts from two different disciplines (linguistics and mathematics), written in English, Ukrainian, and Russian. As a final example, Ozturk's (2007) study and that of Jalilifar (2012b) both examined how the structure of research article introductions varied within linguistics subdisciplines. Not all studies focusing on specific sections of writing samples are based on genre analysis however. Gollin-Kies (2014) utilised a corpus of 352 research articles in ESP, analysing the methods sections so as to ascertain which research methods were more commonly used. Focusing more on specific linguistic features, lexical bundles were analysed in

the introduction, methods, and results sections of 200 applied linguistics research articles (Shahriari, 2017), while Samraj's (2013) study investigated the form and function of citations in the discussion sections of master's theses and research articles in biological sciences. Citation use was also examined by Jalilifar (2012a), again using master's theses and research articles, but this time focusing on the introduction sections.

Text integrity may also be affected by factors other than research design, and it is important to consider the consequences of analysing incomplete texts. Thompson (2006:4) points out a potential disadvantage, citing the case of large, general corpora, in which:

> …texts are often not complete texts, as the corpus design may dictate that lengthy texts are merely sampled; the relevance of this observation is that this makes full text analysis impossible, and thus restricts the range of analyses that can be conducted.

This may be at least part of the reason why many studies do utilise corpora of complete texts. Nonetheless, there is still great variation in the types of text employed for the purpose, and reviewing some of this variation may make clearer why given text types are chosen, and the benefits of these choices for study design.

It is not only the perhaps more obvious genres of academic writing, such as research articles, theses, and coursework essays that have been used in corpora studies in the past. Doctoral students' citation use in grant proposals was examined by Fazel and Shi (2015) in order to see how it compared to that of established academic scholars, while Ding (2007) analysed the discourse and lexical features of 30 medical and dental school application letters. Krishnamurthy & Kozem (2007) also report on other corpora containing text types such as monographs, subject manuals, examination papers, and course packs. It is fair to say however,

that the majority of corpus studies examining the features of academic writing have focused on genres such as theses, dissertations, essays, and research articles.

It is not difficult to understand why research articles are an appealing genre for corpus linguists. They not only represent the predominant arena for the dissemination of disciplinary knowledge, but are also highly conventionalised in terms of academic discourse, and are thus sometimes viewed as models of how to write academically, being presented to students as a 'target' genre (Hyland 2008b). Research articles also reflect the variation in academic writing conventions across disciplines (Bhatia, 2002; Hyland, 2000), which makes them an ideal tool to use in examining these conventions, particularly in contexts where a wide variety of academic disciplines needs to be considered. Additionally, and particularly since the vast majority of academic journals have gone online, electronic research articles are a readily available and convenient source of material for corpora. A large number of studies have utilised the research article genre, either in isolation or as a point of comparison to other academic genres. Studies solely employing research articles include McGrath's (2016) study of self-mentions in 36 history and anthropology research articles; Wang et al.'s (2008) establishment of a medical academic wordlist; Okamura's (2008) study into citation in physics, biology, and chemistry research articles; and Hyland's (1996) investigation of hedging in science research articles. Comparisons of academic language in research articles and student genres can be found in, among others, Cortes (2004), Hyland (2008b), and Jalilifar (2012a). When it comes to student writing itself, there are different justifications for its use, and studies have exhibited substantial variation in the genres that have been analysed and included in corpora.

Nesi & Gardner (2012:28-29) highlight the fact that student genres differ from research genres, in that 'Where research genres aim to persuade the reader of the validity of new findings…assignments generally aim to demonstrate the acquisition of required skills and accepted knowledge'. This, along with the student's status as a 'novice' rather than an 'expert' writer, means that the research purposes of examining student genres may be different from those that motivate the analysis of genres such as research articles. Other than seeking to shed light on the features of student academic discourse in general, studies may focus on the differences between L1 and L2 students, or may focus on pedagogic applications of results. In any event, researchers have a variety of student academic genres from which to choose when compiling or utilising corpora for analysis.

In some instances, longer genres such as the PhD thesis or master's dissertation may be used. An example of this is the study by Petrić (2012), which utilised a corpus of 16 MA dissertations to investigate the use of direct quotation in L2 student writing. The justification given for the choice of this genre was multi-faceted; the substantial length of the theses would ensure that the number of quotes included would be sufficient, the common structure and format would improve comparability, and students would be likely to write to the best of their ability given the high stakes nature of the genre. Hyland also makes use of the master's dissertation, as well as the PhD thesis, to analyse metadiscourse across disciplines (2004), and academic clusters in postgraduate writing (2008b), noting that these, along with research articles, are key academic research genres, and encompass 'the most highly valued kinds of writing produced by students' (Hyland, 2008b:47). Nonetheless, the differences between the two student genres are also highlighted in the latter study. Master's students are required in their writing to demonstrate their intellectual ability, but also to defer to the fact that their readers (lecturers or external

examiners) possess a greater knowledge of the field. In the doctoral thesis on the other hand, the task is to demonstrate sufficient knowledge and expertise through appropriate research and well-constructed argument. As Hyland (2008b:47) notes therefore, 'Assessing the extent of similarities and differences between these genres can offer insights into apprentice and expert performance and feed into classroom practices'. It is presumably reasons similar to these that were the guiding principle leading Charles (2006) to use MPhil and doctoral theses (from social and natural sciences respectively) to examine reporting clauses in citations, given the links made in the study between the investigation of these features and the implications for academic writing pedagogy.

Despite the clear value of theses and dissertations in the study of student academic writing, and the fact that the dissertation is a genre that will be faced by the majority of postgraduate students at master's level, there is also a great deal of other writing done during postgraduate university courses; other forms of summative assessment, such as examinations, often involve essay writing, and assignments given as part of the summative, or indeed formative assessment process may include formats such as research reports, argumentative essays, book reviews, case studies, proposals, and so on. These assignments are not only important in terms of assessment, but also serve to develop students' skills and prepare them for the task of producing a dissertation. It may then, be of considerable value to examine these shorter genres themselves, and a number of previous studies have done so.

Li & Wharton (2012) made use of essays used for assessment in undergraduate courses at universities in China and the UK to examine metadiscourse across contexts and disciplines, while Bolton et al. (2002) looked at connectors in student writing by analysing untimed essays and timed examination scripts. Essays were also the genre employed by Borg (2000) in a study

of citation use in academic writing. Specifically, the corpus was made up of L1 and L2 essays given as a non-assessed first assignment on a master's level TESOL course at a British university. The choice of genre here seems to have been made in order to gain additional insight into L1 and L2 tertiary students' citation behaviours, given a lack of such studies at the time. However, given that the rubric for the assignment specified that no more than one or two references should be used, and that they were, as mentioned, non-assessed, it is perhaps questionable whether these particular essays were the best representation of students' citation behaviour in general. If one wishes to gain an accurate picture of how students write in the university context, examining the kind of writing they are most often required to do would seem to be one useful approach, and in UK higher education this is unlikely to be unassessed writing with minimal references. Similarly, it can be argued that while Hinkel's (2001, 2009) choice of essays written during, or modelled on, university placement tests may have served adequately for the purpose of investigating cohesion and the use of modal verbs in L2 students' academic texts, if one wishes to examine a broader selection of the features of students' academic discourse, more representative genres may be preferable.

Studies that have taken this approach include that of Ädel & Erman (2012), who analysed writing done by students across four terms of study in the case of their L2 sub-corpus, and across two years of study in the L1 sub-corpus, in order to investigate the use of lexical bundles. This would suggest that the essays are likely to have been submitted as components of the students' coursework, and would therefore be an accurate representation of how these students write at university. Assignments produced throughout the academic year as parts of under- and postgraduate students' assessed coursework were also used to compile the BAWE

corpus (Alsop & Nesi, 2009), the University of Warwick Student Writing corpus (Nesi et al., 2004), and the MICUSP (Römer & O'Donnell, 2011).

2.4.2.4 Quality in Corpora

Having selected the source authors, the disciplines, and the genres to be included, there remain still other decisions to be made before the process of compiling a corpus for analysis can be completed. In the case of utilising student writing, one of these decisions is how to deal with the issue of quality. Not all student writing can be regarded as successful in terms of achieving a high, or even proficient grade by whatever marking scheme is applicable. Should a corpus designed to represent student writing include only those examples that have been deemed 'high quality', or should it represent various levels of success? Moreover, is the grade received for a given piece of writing even something that needs to be considered? Can writing that has yet to be assessed be utilised? These are all questions that must be given thought when analysing, and more particularly when compiling corpora.

 In a number of previous studies, the approach has been to restrict the contents of student corpora to writing of high quality. The master's dissertations used in Samraj's (2013:301) study for example, were all 'successful exemplars of the genre', in that within the US university context in which they were produced, students were required to write and re-write their dissertations until they met the required standard. Similarly, the BAWE and MICUSP corpora contain only work awarded high grades (Krishnamurthy & Kosem, 2007). This is perhaps odd, given that, despite only including essays graded II.i (or B+/65%) and above, the BAWE corpus is designed to capture variety in student writing. While Nesi et al. (2005) and Alsop and Nesi (2009) do provide detailed justifications for many of the decisions made in compiling the BAWE

corpus, the matter of why only high-graded assignments were included is not addressed in any detail.

Petrić adopted an alternative approach in her studies of citation and direct quotation (2007, 2012 respectively), including both high and low scoring master's theses in the analysis, since the aim, in part, was to see how grade is affected by the use of these features. Nonetheless, the value of incorporating writing that varies in terms of its academic success is also noted by Krishnamurthy & Kosem (2007:367), who argue that student comparison and the monitoring of progression are impossible without the inclusion of lower-grade texts. Furthermore, they contend, the writers of these texts are likely to be those who are most in need of help in terms of EAP instruction, and that low-scoring texts serve an important role in helping to diagnose and assess problems, and 'address wider issues such as the EAP course content, the teaching materials, and the teaching and learning strategies'.

If the aim of a study is to examine certain features of academic discourse across a wide sample of student writing, as it is here, then the issue of grading may be a problematic one. Petrić (2012) notes the difficulties involved in obtaining writing representative of all score bands when compiling a corpus of MA dissertations, as grades were skewed towards the higher scores. Furthermore, the value of considering how well or badly a piece of writing has been scored is surely in attempting to represent various levels of writing quality in a corpus, but how can we be sure of the extent to which the scores achieved by a particular example of student writing represent the quality of the writing itself? They may be based more on content, as this may be the primary concern of lecturers when marking (Evans and Morrison, 2010). Despite working to detailed marking schemes, individual markers may still differ in the importance they attach to errors, appropriate lexis, or adherence to the academic discourse conventions of the

discipline. Therefore, given the potential impracticality of attempting to establish each individual marker's scoring priorities across, for example, multiple departments of a large university, it may be preferable not to include scores as a consideration when compiling a corpus. If the aim is to build as representative a sample of student writing as realistically possible, then taking a 'snapshot' of assignments being submitted in various subjects and disciplines, irrespective of scores achieved, would seem one way to achieve that. Finally, in purely practical terms, students may be more willing to participate and consent to the use of their writing in a study if they can be assured that grades and feedback will not be considered in the analysis.

Another issue that arises in terms of the quality and authenticity of student writing, particularly in the case of L2 students, is that of proofreading. Many students for whom English is not a first language seek L1 speakers to proofread their work before submission. Does this pose an obstacle to the writing being an authentic representation of how the student writes? Arguably, this depends entirely on which aspects or features of the writing are under investigation. Certainly if a study aims to examine errors for example, then utilising writing that has been proofread would be ill-advised. On the other hand, whereas errors in spelling, lexis, punctuation, and syntax would probably be corrected as a matter of course, the general use of the features of interest in this study - academic vocabulary, lexical bundles, hedging, and citation - particularly in terms of variety or rhetorical function - may be less likely to be significantly amended during the process of proofreading.

In any event, it is common for L2 students to seek the services of proofreaders before submitting assignments, and this is something that is understood and may even be encouraged by academic staff. If L2 (and even in some instances, L1) students regularly submit writing that

has been proofread, then this work must be regarded as legitimately representative of student writing in university, and as long as the academic discourse features of interest are not likely to have been significantly changed or removed, the process of proofreading need not present great concern to the corpus linguist investigating student academic discourse. Moreover, published academic writing too may well have been revised and edited before being accepted for publication, and therefore it seems reasonable to accept the proofing process as an element in producing academic writing, be that for publication, or submission in a higher education institution.

As outlined in the preceding sections, there are many considerations when compiling a corpus of academic writing, including the source writers and how they are compared, the variety of disciplines included, the genres of academic writing from which the corpus will be built, and the issue of how to approach the quality of any student writing utilised. The decisions made with regard to these considerations must take into account the anticipated purpose of the corpus, and if it is to be constructed for a specific study, then the specific research aims and questions it is designed to help answer. These will necessarily affect the design of the corpus, and the methodology employed to interrogate it.

Having reviewed the relevant literature and provided a theoretical background to the study that helps to situate it within the wider research context, the following chapter details the study's methodology, including the research paradigm, the overall design, and the specific methods employed to prepare and carry out data collection and analysis.

3 Methodology

3.1 Overview of Chapter

This chapter details the methodology behind the study and the procedures carried out to conduct it. This includes all relevant ethical issues, and the research paradigm and overall research design, followed by a specific, comprehensive description of how the study was carried out, including the procedures for both data collection and analysis.

3.2 Ethical Considerations

Prior to commencing the research itself, full ethical approval was obtained. This included the institution's standard ethical checklist and insurance documentation, approval of ethical amendments, legal confidentiality agreements, and an Extraction Plan with reference to accessing the online Moodle/Turnitin submission system. Prior written consent was obtained from all study participants (See Appendix A for all relevant ethical documentation).

3.3 Research Paradigm

Before describing the research itself, it may be useful to consider the study's underlying research paradigm, since it has been suggested that an understanding of the philosophical underpinnings leading to the choice of research questions and methodology is important if research is to be conducted with precision and clarity (Grix, 2004). This must be considered in light of ontological and epistemological assumptions. Put succinctly, ontology is 'one's view of reality and being' while epistemology refers to 'the view of how one acquires knowledge' (Mack, 2010:5). Ontology therefore concerns the nature of existence, and raises the question of whether reality is objective or shaped by the observer's interpretation. Epistemology on the other hand concerns the nature of knowledge and how the inquirer and the known are connected, forcing us to ask whether knowledge is 'a set of value-free truths, transcending

opinion, or whether it is subjectively built and progressively understood, through experience'
(Brooke, 2013:430). The answers to these questions will impact upon the research paradigm
and methodology of a study.

Research paradigms can be broadly categorised on a continuum between interpretivist and
positivist, and these positions make differing assumptions in terms of ontology and
epistemology. While interpretivists assume that there is a connection between reality and the
individual observing it, and that knowledge is a reflection of their own history, culture and
experience, the positivist paradigm assumes a separation between reality and the observer,
with human knowledge being based on an independent, objective reality (Weber, 2004).
Accordingly, these paradigms will tend to lead to different approaches when it comes to
research methodology. The interpretivist paradigm favours a qualitative approach to research,
emphasising a flexible, context-sensitive framework, through which we can understand the
meaning of social experience from the perspective of the participants. Conversely, research
adopting the positivist paradigm relies heavily on quantitative methods reflecting the scientific
approach, measuring and analysing causal relationships within a logical, value-free framework
(Brooke, 2013; Mack, 2010; Yilmaz, 2013).

In focusing predominantly on the use of CL, the present study grounds itself largely within the
positivist paradigm. In their discussion of the nature of corpus linguistics, McEnery and
Gabrielatos (2006) contend that CL is empirical, 'in that it examines, and draws conclusions
from, attested language use, rather than intuitions' (p.34). They go on to state that while CL
does not completely reject the idea of data being interpreted qualitatively, it nonetheless
focuses heavily on quantitative information. The methodology of this study must therefore be
regarded as predominantly quantitative in nature, and therefore also as fitting within the

positivist research paradigm. That said, combining qualitative and quantitative methods can result in a research design that enhances the strengths of each method while mitigating the weaknesses, as well as improving general validity; 'words can add meaning to numbers, and numbers can add precision to words' (Dörnyei, 2007:45). With reference to CL in particular, 'linguists who neglect corpora do so at their peril, but so do those who limit themselves to corpora' (Johansson, 1991:6). This study therefore utilises a mixed-methods approach, incorporating as an addition to the corpus analysis, a limited qualitative element in the form of semi-structured interviews.

3.4 Overall Research Design

This study uses corpus analysis to usefully inform pedagogy for writing in a university pre-sessional English course. Since no existing corpus was considered to have the specificity of design necessary to facilitate the study, it was decided that a corpus would be built for, and as an element of the research. The details of how this was achieved, and the design decisions that were involved in this process are discussed in section 3.4.

 For reasons of both theory and practicality, the analysis itself focuses on a limited selection of academic writing features identified through reviewing previous studies, particularly those utilising CL. Details of the analysis and the rationale behind the decisions made in its formulation are provided in Section 3.6.

 In order to provide triangulation for the data obtained from the corpus analysis, and to complement the objective linguistic data with a more subjective element, taking into account the thoughts and opinions of learners, the second section of the study consists of qualitative, semi-structured interviews. These were conducted with students currently enrolled at the institution, who had completed at least one term of their programmes of study, and who had

previously completed one of the institution's pre-sessional courses (5, 10, or 20 weeks in duration). Section 3.4 details this element of the study.

Finally, the course content and materials of the institution's existing pre-sessional course were examined, with specific reference to the writing component of the course, so as to provide a frame of reference against which to view the findings of the corpus analysis and interviews, and from which to assess the potential of the methodology to usefully inform course content in this context.

3.5 Creation of Corpora

In light of the considerations detailed in sections 2.4.2.1 – 2.4.2.2, the corpus compiled for this study aimed to include L1 and L2 PG student work, as well as comparable published sources; writing samples from a wide range of the academic disciplines represented at the institution; and assignments submitted as part of the formative or summative assessment ongoing throughout the academic year, irrespective of 'quality' in terms of being scored for that assessment.

The inclusion of L1 student, L2 student, and published sources offered a means of triangulation where the corpus was concerned. If an analysis of L2 student writing is to inform pedagogy and/or teaching materials, there must clearly be a point of comparison – some 'standard' model of academic English against which to examine L2 student writing. Previous studies have, for the most part, employed two-way comparisons, either of L1 and L2 student writing (Borg, 2000; Friginal et al, 2014; Granger & Tyson, 1996; Hinkel, 2001; Hyland & Milton, 1997) or of L2 student and 'expert' sources (Hyland, 2008b; Jalilifar, 2012, for example). While these approaches can certainly provide useful insights, both have also been criticised. In the case of L1 student writers, they too are often 'novices' when it comes to academic writing

conventions, and may therefore not represent ideal models against which to judge the efforts of L2 writers. The published writing of 'experts', while arguably a more reliable model in terms of accuracy and academic discourse community conventions, differs from L2 student writing in terms of its intended purpose and audience, so may also not be an ideal point of comparison. This study sought to mitigate some of these issues by means of a three-way analysis, comparing L2 student writing against that of both L1 students *and* published sources. This approach had a number of advantages. Firstly, it removed the restrictive framework in which L1 student or published academic writing are regarded individually as 'standard' models. By analysing the three writing sources against each other, it was possible to establish not only how L2 student writing differed, but also the areas in which the writing of L1 students and published writers showed variance. This allowed any inferences drawn to be more fully informed; rather than simply being based on an L1/L2 or student/expert dichotomy, the findings of the corpus analysis could be considered in light of the three-way comparison. For example, if the usage of a given feature was similar among L1 students and published writers, but differed in the case of L2 students, this might suggest that linguistic proficiency was a likely cause, whereas if a given feature was used in much the same way by both student groups, but showed marked differences among the published writers, the notion that such variance might be a product of the different genres (coursework vs. research article) could be considered. Secondly, by highlighting not only L2/L1 student and L2 student/published usage differences, but also L1 student/published variance, the study was able to throw light on potential issues faced by *both* student groups. Given that, as mentioned above, L1 students are often also novice academic writers, gaining insights into those areas in which they may benefit from additional instruction is also valuable, particularly when considering the optimal future models

for university EAP/writing instruction. While conducting a three-way comparison undoubtedly presented additional challenges in terms of the size and scope of the data, and the number of analyses that needed to be carried out, the advantages of this more rigorous and informative approach outweighed any data management issues that arose.

Since the importance of variation across academic disciplines has been highlighted numerous times (Ädel and Erman, 2012; Harwood, 2005; Hyland, 2002a; Hu, 2007; Paltridge, 2002), it was felt that the corpus should reflect this variation as far as possible, especially since this was one of the aspects of pre-sessional content that the study aimed to consider. The extent of the disciplinary variation reflected in previous corpora of academic writing has varied. For example, the MICUSP (Römer & O'Donnell 2011) contains 16 disciplines, the corpus created by Hyland and Tse (2007) includes work from eight disciplines, Cao and Xiao's 2013 study of L1 and L2 writer abstracts included 12 academic disciplines, and the BAWE corpus represents over 30 different disciplines (Alsop and Nesi, 2009).

For the present study, the intention was to incorporate 15 disciplines in total, representing three from each of the five faculties at the institution. The selection of specific Level 7 courses representing these disciplines was made so as to include those which had relatively high levels of international student enrolment in the 2017-18 and/or 2018-19 academic year, based on figures supplied by the institution. 'International student' is used here to incorporate all non-UK domicile students. This is an important distinction from the generally understood definition; UK HEIs classify students into three groups – Home (UK) students, who are ordinarily resident in the UK; EU students, who are nationals of EU countries and are ordinarily resident in the EEA (European Economic Area) or Switzerland; and international students, who are those not falling into either of the above categories. In reference to the selection of courses in this study,

the term 'international student' here includes not only those whom the HEI sector would define as international, but also those it would define as EU students. Selection based on international student enrolment was considered the most efficient way of identifying courses likely to have higher numbers of L2 students. Of course, there may be cases where international students are L1 English speakers, or where UK domicile students are L2s, but this could not practically be addressed within the course selection process, and no preferable, practicable method of identifying courses for inclusion in the study existed. The L1/L2 status of each individual student contributing to the corpus was established when they consented to participate in the study. The 15 disciplines initially chosen for inclusion in the study are shown in Table 1.

Table 1-Initial selection of disciplines based on international student enrolment numbers

Faculty	Subject Area
Arts and Humanities	Architecture
	TESOL
	Fashion
Education	PGCE Languages
	Inclusive Education
	Educational Leadership
Business and Law	Professional Bar Training
	Advertising
	Marketing
Health Psychology and Social Care	Social Work
	Forensic Psychology
	Physiotherapy
Science and Engineering	Geographical Information Systems
	Data Science
	Cell Biology

3.5.1 Corpus of Students' Academic Writing

The corpus of students' academic writing created for the present study was compiled from formative and summative assessment submissions made by postgraduate students at the institution during their courses of study. Coursework submissions were chosen as the focus of analysis rather than other genres of student writing such as theses or dissertations. These had a number of practical advantages: they were potentially relatively higher in number, since they were not restricted to one per student; in terms of data collection they became available at various stages during the academic year rather than being submitted after the completion of the summer term, thus making data collection less problematic given the time constraints of the study; and they could encompass a variety of text types, thus providing a more representative picture of student academic writing in general than would the more restrictive genre of theses and dissertations. As mentioned in Section 2.4.2.3, there is also a research precedent for the use of coursework submissions, given that these were utilised in the compilation of the BAWE, MICUSP, and University of Warwick Student Writing Corpus (Alsop & Nesi, 2009; Nesi et al., 2004; Römer & O'Donnell, 2011). Once the courses had been provisionally selected (see above), it was necessary to proceed through a number of different administrative stages before reaching a point where data could realistically be gathered and the corpus could be built. Initially, the programme leader for each course was identified, and approached by email with an outline of the study so as to establish their willingness, in theory, for the course and its students to be included in the study. At this stage, and for issues of practicality, a number of programme leaders were not able to provide the necessary access to their courses, or judged them unsuitable for the study's purposes. This unfortunately included

all three courses from the Science and Engineering Faculty, as well as the Marketing subject, and it was therefore not possible to include these courses.

Once a programme leader agreed to the inclusion of their course, I liaised with them and other faculty teaching staff in order to arrange an opportunity to secure individual consents from students to access and utilise their written coursework submissions. In most cases, this took place during lectures or tutorials in the autumn and spring terms of 2018/19, although in the cases of Architecture, Bar Training, and TESOL, it was also possible to secure a number of completed consent forms from the 2017/18 student cohort during the final term of that academic year. The study was explained to the students, and they were each supplied with a Participant Information Sheet (PIS) and a consent form (see Appendix A) These forms granted me permission to access any submissions made since the student's course of study began, and any made in the future, allowing collection of work submitted prior to and after the completion of the consent forms themselves, and thus potentially maximising the amount of student work that could be incorporated into the corpus. As Hardy & Friginal (2016) observe, methodologies based on the analysis of large corpora can allow researchers to make generalized observations with more validity than would be possible with smaller samples. The consent form was also used to establish, through means of a tick box, whether the individual identified themselves as an L1 or L2 English speaker. It is important to note here that while the terms L1/L2 are used throughout this study so as to avoid the use of 'native speaker' (NS) and 'non-native speaker' (NNS) (terms which are potentially problematic and raise issues related to the notions of native-speakerism, ethnocentrism, and the idea that this classification automatically suggests NS superiority), the NS and NNS labels *were* employed for the purposes of the consent form, as it was judged that these would be the most widely understood terms, particularly for those

with little or no background in language education. While it may represent a false dichotomy to categorise English users on the basis of whether English is or isn't their first language, the practicalities of the study were such that it was not possible to establish in detail the particular language backgrounds or proficiencies of all those individuals contributing to the corpus, and as such, some manner of relatively straightforward labelling system was required to discriminate between English users – even if that system was not ideal. Paper consent forms were stored securely, and scanned versions were stored on a secure laptop and encrypted hard drive.

Having secured student consents, the next stage in the process was to acquire access to the written submissions themselves. Many courses across the institution utilise the Turnitin system for electronic submission, and it was judged that in terms of practicality, this would provide an ideal, centralised means of accessing the students' written work. This approach would also minimise the disruption and inconvenience caused to faculty staff and students, by avoiding a scenario in which they were required to actively supply written submissions to the study.

The Moodle interface for the Turnitin submission system is accessed by means of log-in with an institution ID number. Staff and students are granted distinct levels of access, pertinent to their needs. In order to access student submissions from the various different courses, it was necessary to secure staff-level access to the Moodle system, and specifically to each individual relevant unit for each course. A guest staff ID was generated by the institution's IT department for this purpose.

Due to the large amount of confidential information stored within the system, there were a number of data protection issues that also needed to be addressed if I was to be granted this

staff-level access. Through liaison with the institution's legal office, a confidentiality agreement (see Appendix A) was drawn up, detailing the limits of my access and my responsibilities in terms of ensuring both that these limits were adhered to, and that the confidentiality and security of any information accessed was ensured. This document was then supplied to the programme leaders prior to requesting that I be granted access to their subject units on the Moodle system. As part of the arrangements made with the legal office, an extraction plan was drawn up, re-affirming my commitment to the data protection process. This also required that for any submission downloaded from the Moodle system, the time and date would be recorded, along with the name and institution ID of the individual student by whom the submission had been made. These details would then be passed to my department head, so that this record of my activity within the system could, if necessary, be compared to that logged automatically within the system itself. This would provide additional accountability. With these confidentiality and data protection measures in place, I was able to gain access to the Moodle submission system with the assistance of the programme leaders in each subject area.

Once access was given, it was possible to view and download the coursework submissions (in their originally submitted form, i.e. free from tutor feedback, comments, corrections or grades) from those students who had consented to the use of their work. The decision to utilise submissions irrespective of grade or 'quality' was based on the considerations previously outlined in Section 2.4.2.4, and on a variety of practical concerns. It was unclear exactly how much student writing it would be possible to obtain, and excluding submissions based on grade would have restricted the potential size of the corpus further. Moreover, it was felt that the students themselves would be more likely to consent to their work being used if they could be

reassured that grades and feedback would in no way be included or considered as part of the analysis. Finally, in light of the data protection and consent considerations, including grades and/or feedback may have resulted in the necessity of obtaining additional consents from the staff involved in the marking process, something that would have been impractical given the timeframe available for collection of submissions.

Each submission, once downloaded, was prepared for the corpus by removing title pages, contents pages, all tables and illustrations, headings and subheadings, bibliographies and appendices. Additionally, any excerpts from transcriptions of interviews were also removed, as these would represent spoken English, rather than the academic writing of the students themselves. All files were then converted to ASCII plain text format and named using a system that protected the anonymity of the student writers while still allowing me to trace the authorship if necessary (for example in the event of an individual student withdrawing from the study and their written contributions having to be removed). Filenames also served to identify the main subject area, and in cases where submissions from multiple units within a subject had been collected, were numbered so as to identify the unit specifically. As an example, a file from a student 'John Smith' studying TESOL, might be named 'Tesjosm1', utilising a prefix to identify the subject, and then the first two letters of the forename and surname, along with a number to identify the unit from which the submission came.

While, as mentioned, the majority of courses at the institution utilise the electronic submission system, there are still some that do not. This included the areas of Architecture and Bar Training. For these courses, it was necessary to rely on the students themselves to provide their coursework via email. This proved not to be an effective means of gathering the necessary written work, and with only minimal submissions being made available, it was

regrettably necessary to exclude these courses from the study. Additionally, two of the courses from the Faculty of Education, in the areas of Inclusive Education and Educational Leadership were merged for the purposes of the study, into a single 'Education' discipline. This decision was taken as a result of the limited number of participants (n=10 and n=5 respectively), which would have led to a very low number of submissions for analysis. While it may seem that merging these courses was somewhat arbitrary, there are units common to both courses, such that a legitimate connection can be made between them, and it was felt that this was preferable to excluding them from the study.

Once the consent forms had all been collected, and the decisions had been made on the inclusion/exclusion of the various courses, the final result was that the study would focus on eight subject areas from four of the institution's faculties, as shown in Table 2.

Table 2-Final selection of disciplines

Faculty	Subject Area
Arts and Humanities	TESOL
	Fashion
Education	PGCE Languages
	Education
Business and Law	Advertising
Health Psychology and Social Care	Social Work
	Forensic Psychology
	Physiotherapy

Subject-level labels were used to represent disciplines, rather than, for example, faculty-level categories. It was felt that this would provide a clearer and more accurate representation of the disciplines included in the corpus, since they could not be regarded as representing the

overall range of disciplines in a given faculty or broader academic field, and as such, utilising wider, more inclusive field or faculty-level categories could potentially have been a misrepresentation of the contents of the corpus.

3.5.2 Corpus of Published Academic Writing

Academic staff and/or the institution's specialist subject librarians in each of the chosen disciplines were consulted with regard to potential sources for the corpus of published writing. They were asked to suggest publications which they considered to represent 'model' writing in their particular discipline, or publications to which students were frequently referred and which they may therefore view as exemplars of writing in their area of study.

These recommendations centred around academic journals, and although there has been much debate over whether these represent a suitable source of comparison for student academic writing (Gardner and Nesi, 2012; Hyland, 2008b; Jalilifar, 2012b), there are, as discussed in Section 2.4.2.3, a number of advantages to their use. These include their conventionality, their frequent presentation as a target genre, their reflection of disciplinary conventions, and in practical terms, ease of access. Additionally, in terms of facilitating comparison, research articles have the advantage of being complete texts of a length not dissimilar to that which students may be required to produce in assignments. For this study in particular, given that one of the main features of interest was citation, research articles also represented an ideal source, since other published academic genres such as textbooks for example, do not necessarily contain in-text citations, and would thus not be useful comparators in this regard. In light of these considerations, and the fact that subject staff themselves had recommended research articles as appropriate sources of comparison to their

students' academic writing, it was felt that utilising this resource would be a legitimate approach for the present study.

 In each of the eight subject areas, three journals were chosen based on staff recommendations. From each journal, three research articles per issue were randomly selected by assigning each article a number and using a random number generator. This process began with the most recent issue at time of compilation, and moved back through the previous four issues, so that a total of 15 articles from the five most recent issues of each journal were collected. As with submissions for the student corpus, all articles were processed so as to remove administrative information such as author contact details and submission dates, as well as titles, tables, figures, acknowledgments and end notes (where present), verbatim interview extracts, and reference lists. Files were then converted to ASCII plain text format, suitable for analysis, and named so as to allow the original source material to be identified. The final corpus of published writing consisted of 360 research articles. The full list of journal publications utilised in the study is given in Table 3.

Table 3-Journal publications utilised, by subject area

Subject Area	Journals
Advertising	International Journal of Advertising Journal of Consumer Marketing Journal of Marketing Communications
Education	International Journal of Inclusive Education European Journal of Special Needs Education Cambridge Journal of Education
Fashion Merchandising	Journal of Fashion Marketing and Management Journal of Management Studies Journal of Brand Strategy
Forensic Psychology	Journal of Forensic Practice Journal of Forensic Psychiatry and Psychology Legal and Criminological Psychology
PGCE Languages	Language Learning Journal Language and Education Educational Research
Physiotherapy	Physiotherapy Physical Therapy British Journal of Sports Medicine
Social Work	British Journal of Social Work Child and Adolescent Social Work Journal Journal of Social Work
TESOL	Applied Linguistics TESOL Quarterly Modern Language Journal

3.5.3 Completed Corpus

The details of the final corpus are given in Table 4, below.

Table 4-Word counts for corpus, by discipline and writer group

Discipline / Writer Group	Advertising	Education	Fashion	Forensic Psychology	Languages	Physiotherapy	Social Work	TESOL
L2 Student	49696	88678	53246	38812	102370	6590	19033	262512
L1 Student	140622	49870	150677	276292	82230	22492	80168	338100
Published	302083	265841	303978	218408	287310	167298	267995	366970
Total	492401	404389	507901	533512	471910	196380	367196	967582

In terms of writer groups, the final corpus resulted in a count of **620,937** words for L2 student writing, **1,140,451** words for L1 student writing (a combined total of **1,761,388** for students as a whole), and **2,179,883** words of published writing. The entire corpus consisted of **3,941,271** words. As can be seen from the table above, there is considerable variation in the wordcounts across disciplines and writer groups. This was unfortunately unavoidable given the writing sources and the numbers of students participating in each discipline. The results will be normalised as far as possible to take account of these differences, but in the case of physiotherapy (the L2 student sub-corpus in particular), any inferences must necessarily be made with caution, and generalisation avoided.

3.6 Semi-Structured Interviews

This section of the study was carried out in order to complement the data from the corpus analysis with more qualitative, subjective insights from L2 students who had not only completed one of the institution's pre-sessional courses, but had also spent at least one full term studying their level 7 course at the institution. The participants would potentially be in a

position to consider their experiences with academic writing in English and relate this to their experience of the pre-sessional course. This phase of the study therefore aimed to establish a) the challenges participants experience in academic writing, as well as their awareness of the various academic writing features under consideration in the study, and b) how they perceive the pre-sessional course to have benefitted them in terms of their academic writing ability. This would demonstrate the extent to which the findings of the corpus analysis highlighted areas also mentioned by the students themselves. Moreover, 'It is important and appropriate to question students on their sense of what progress they make and what difficulties they encounter as they move from our writing classes to other writing demands in content courses' (Leki and Carson, 1994:95).

 A semi-structured interview format was chosen, so as to provide some flexibility in the question schedule, allow points of interest to be pursued in more depth, and give interviewees the opportunity to elaborate upon those areas in which they might have more to contribute (Alsaawi, 2014; Qu & Dumay, 2011).

3.6.1 Interview Participants

In August 2018, I volunteered at the institution's extra-curricular conversation club for current pre-sessional students, and was thus able to introduce my research and appeal for potential participants. Eighteen pre-sessional students supplied contact details and provisionally agreed to participate in interviews to be carried out in early 2019, after completion of the first term of the 2018/19 academic year. These 18 students were re-contacted via email in February 2019, and asked if they were still willing to participate in either a Skype or face to face interview. Of the 18 students contacted, 6 responded to the request for interview.

All participants were L2 English speakers who had completed either the 20, 10, or 5 week pre-sessional course, and were now engaged in studying Level 7 courses at the institution (see Table 5, below)

Table 5-Details of pre-sessional courses and Level 7 subjects undertaken by interview participants

Participant	Pre-sessional course	Level 7 Programme
P1	5 week	Product Design
P2	10 week	Architecture
P3	5 week	Architecture
P4	20 week	Banking and Finance
P5	5 week	Architecture
P6	5 week	Architecture

Interviews were conducted individually, face to face, on the university campus. Participants were first given the Participant Information Sheet, and were asked to sign the Consent Form, stating that they agreed to participate in the study, they acknowledged that interviews would be recorded, and that they understood that they could withdraw their participation at any time. As per the recommendations of Robson (2011), interviews began by me introducing myself once more to the participants, reminding them of the interview focus, and asking a few simple questions to put them at ease. These first questions established basic information such as which pre-sessional course they had taken and their current course of study. Participants were then asked about the challenges they perceive themselves to face in writing academically in English, about their experiences with writing for their Level 7 courses, and about their confidence with reference to understanding and using particular aspects of English academic writing, namely academic vocabulary, lexical bundles, hedging, and citation.

Dörnyei (2007) stresses the importance of piloting open-ended questions such as those commonly used in semi-structured interviews; unfortunately however, due to time constraints and limited access to participants, pilot interviews were not possible in this study. In order to mitigate for the absence of a pilot study, the precise wording of the interview questions was honed between each interview, so as to adapt to any apparent problems of comprehension on the part of the interviewees. It is important to note however, that any alterations made to the wording of questions were carefully considered in order to ensure that the overall meaning of the questions was maintained throughout the process. Thus, while the first and last participants for example may not have been asked identical questions in terms of wording, the meaning of the questions, and the information requested of the participants was consistent. This careful process of adaptation allowed any issues to be addressed while still maintaining the integrity of the interview question schedule, and was the optimal approach in the absence of a pilot study. It was important to ensure that interview questions were phrased in such a way as to make clear the concepts that were being discussed, without unnecessarily feeding information to participants through the questions themselves. This was because the aim was to establish how familiar the participants were with ideas such as hedging and citation practices, and because their experiences and recollections relating to the teaching of the various writing features on their pre-sessional courses was also of interest. In addition to the specific features of academic writing, participants were also asked about their general impressions of their pre-sessional course in terms of how it had prepared them in respect to academic writing, and about any aspects of the course they had found particularly useful (or otherwise) in this regard (see Appendix B for the outline interview schedule).

It is important to bear in mind that some of the areas of enquiry in the interviews relied on participants' recollections of courses they had taken some months earlier. While interviewees were not requested to recollect their pre-sessional course content in any specific detail, the interpretation of their responses must nonetheless be carried out with an awareness that, for example, simply because they may not have recalled something being included on the course, does not necessarily mean it was not.

Unfortunately, given the small number of interviewees, this section of the study offered less utility than had been hoped, and while the potential for useful qualitative insights remains, it is important to consider the limitations of any findings, particularly in terms of their generalisability. For this reason, data from the interviews will be used only in a limited capacity, for the purpose of providing tentative qualitative insights in conjunction with the data from the corpus analysis.

3.7 Examination of Pre-sessional Materials

Following requests to staff involved in the course, in the 2018/2019 academic year I was supplied with PDF versions of all writing-related materials used on pre-sessional English courses at the institution[1]. These materials covered the five-week course, the ten-week course, and the final ten weeks of the 20-week course. During the first ten weeks of the 20 week course, students use the Oxford EAP Intermediate textbook (De Chazal & Rogers, 2013), so this was also included in the examination. The materials were searched manually for any reference in writing materials to the four features important to the study, both as direct teaching points

---

[1] Materials were confirmed as current and still in use as of February 2020.

and when included as part of exercises/activities in which they were not the primary teaching focus.

3.8 Data Analysis

3.8.1 Vocabulary

Following the methodology of Hyland and Tse (2007), which investigated the distribution of AWL families in an academic corpus, and was therefore based on similar analyses to those required in this study, the Range32 software package (Heatley and Nation, 2002) was used to analyse the vocabulary use within the corpus. Range allows a set of texts to be compared against base lists of words, so that the coverage provided by the words in each list can be ascertained. The software includes three base lists: the first and second 1000 most frequent English words from the General Service List of English Words, or GSL (West, 1953), and the AWL (Coxhead, 2000). Range output gives information on how many of the tokens and types in the analysed texts come from each of the base lists, and also expresses this as a percentage of all the running words in the text. This means that the percentage of the vocabulary in the texts covered by the first 2000 words of the GSL and by the AWL can be established. Additionally, the number of word families in the texts originating from each base list is also given, along with the number and percentage of types and tokens that do not fall within any of the three lists. Finally, the software provides lists of the word families in each base list, with the frequency of occurrence of each item.

While other lists of academic vocabulary have been created since the AWL, including the New Academic Word List (NAWL) (Browne et al., 2013), and the Academic Vocabulary List (AVL) (Gardner & Davies, 2014), the original AWL was used in this study for a number of reasons. Firstly, 'the AWL is still a useful tool for EAP instructors who teach students from a variety of

academic disciplines in general EAP courses rather than discipline specific English for Specific Purposes (ESP) courses' (McDonough et al., 2018:77) and is still widely used in research. Secondly, although other software packages, such as AntWordProfiler (Anthony, 2014) and Web Vocab Profiler (Cobb, n.d) are available and can be used for analysis with some of the newer lists, issues such as document size, ease of data transfer, and the large number of analyses necessary, meant that Range32 (which uses the original AWL and GSL lists) was the best choice in practical terms. Finally, while the more recent lists have advantages such as being compiled from larger corpora and including more academic disciplines (Higginbotham & Reid, 2019), and may therefore provide a more accurate baseline of what represents academic vocabulary, the aim of this study is not primarily to assess the extent of academic vocabulary employed in the corpus. Rather, the study aims to examine the degree to which academic vocabulary usage varies between writer groups and disciplines, and as such the AWL provides a suitable measure for these purposes.

The three sub-corpora, L2 student, L1 student, and published (hereafter L2S, L1S, Published) in each of the disciplines were compared individually against the three base lists, in order to establish how AWL coverage in the three writer groups differed. In most cases this could be achieved by simply selecting all the plain text files comprising a given discipline sub-corpus, and importing them into Range for analysis. However, since the software has an upper limit of 60 texts per analysis, the TESOL L1S and L2S, and the Forensic Psychology L1S sub-corpora (containing 107, 82, and 100 texts respectively) had first to be merged into single text files before they could be imported to Range. In addition, specific issues were encountered with respect to the vocabulary analysis, details of which are reported in Section 4.1.

3.8.2 Lexical Bundles

For this study, 4-word bundles were chosen for the analysis. The choice of 4, rather than 3 or 5-word bundles was based on Hyland's (2008b) rationale, which takes account of commonality of occurrence and range of structure and function. Four-word bundles are also 'frequent enough for good diversity' (Bestgen, 2018:206). Previous studies have differed in terms of the cut-off points used to determine the inclusion or exclusion of bundles produced in cluster lists. Two criteria are usually considered for this: frequency and dispersion. The former measure is designed to 'identify bundles that recur often enough to be regarded as typical of the target register' (Pan et al., 2016:63), while the latter avoids 'idiosyncratic uses from individual writers' and ensures that bundles are more representative of the corpus, rather than being restricted to a few writers or texts only (Bychkovska & Lee, 2017:42; Pan et al., 2016).

In terms of frequency, for large written corpora the normalised cut-off tends to be somewhere in the region of 20 to 40 occurrences per million words (Chen & Baker, 2010), although the determination of this cut-off can be somewhat arbitrary (Biber et al., 2004). Biber et al. (1999) for example, used a cut-off frequency of ten times per million words, while Cortes (2004), Hyland (2008b), and Liu (2012) employed the more conservative cut-off of 20 times per million words, and Biber et al. (2004), Bychkovska & Lee (2017), Esfandiari and Barbary (2017), and Pan et al. (2016) chose a threshold frequency of 40 occurrences per million words. This study has also used the 40 times per million words cut-off, on the basis that it has become the standard in recent lexical bundle studies (Bychkovska & Lee, 2017; Pan et al., 2016), and has been demonstrated to represent an appropriate cut-off for four-word bundles (Bestgen, 2018).

Cut-offs for dispersion also exhibit some variation, with thresholds of 10% of sample texts (Hyland, 2008b), 5 texts (Biber et al., 2004; Bychkovska & Lee, 2017; Cortes, 2004; Esfandiari & Barbary, 2017), and 3 texts (Chen & Baker, 2010) being utilised in previous studies. The intention in the present study was to adopt the threshold of bundles occurring in 5 or more texts; however, due to the small number of texts in some of the sub-corpora, this value would have resulted in very limited results or would even have removed the possibility of identifying any bundles at all in the case of the physiotherapy L2S sub-corpus, which contains only 3 texts in total. It was therefore decided that the search criteria would be expanded to include bundles present in 3 or more texts within each sub-corpus. It was anticipated that this would not skew results across the rest of the sub-corpora, as in practice the dispersion restriction 'has little effect, because most bundles are distributed widely across the texts in a corpus' (Biber et al, 2004:376). It has been noted however, that despite normalisation such as that outlined above, smaller corpora may generate more bundles than larger corpora, and even bundles that repeat only a small number of times can meet the cut-off requirements (Cortes, 2015; Hyland & Jiang, 2018). This means that in cases where sub-corpora of notably different sizes are being analysed, there is a potential for results to be skewed. For this reason, following the application of the cut-off requirements for frequency and dispersion, only the 20 most frequently occurring bundles resulting from each search were considered in this study.

For each discipline, lexical bundles within L1S, L2S and Published were examined by first generating index files in the Wordlist utility of Wordsmith Tools (Scott, 2019), and then using these to produce cluster lists. Bundles that did not meet the criteria for inclusion in terms of frequency or dispersion were then excluded. Lists were produced:

- for the entire corpus (all disciplines and all sub-corpora combined)

- for each writer group in each discipline (to compare L1S, L2S and Published within and across disciplines).

3.8.3 Hedging

It is important primarily to define what qualifies as a hedge for the purposes of this study, since no precise and universally agreed-upon definition of a hedge exists (Vold, 2006). Hyland (1998:5) offers the following definition: 'the means by which writers can present a proposition as an opinion rather than a fact', while Crompton (1997:281) opts for a narrower outlook, arguing that a hedge is, 'an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters.' This is similar to Lyon's (1977:797) definition of epistemic modality – '[a]ny utterance in which the speaker explicitly qualifies his commitment to the truth of the proposition expressed by the sentence he utters', and indeed Crompton goes on the argue that there is a strong case for tying hedging to epistemic modality. Various studies have used epistemic modality markers to investigate hedging. These markers, arguably 'a dominant and basic type of hedge' can be described as 'linguistic expressions that qualify the truth value of a propositional content', (Vold, 2006:65). They include those elements that qualify the commitment (boosters) or lack of commitment (hedges) on the part of a writer to the truth of a proposition (Vasquez & Giner, 2008). This study will focus on lexical hedges. While this may exclude other means by which a proposition could be hedged (such as through voice or tense for example), lexical items are commonly associated with hedging, and represent the predominant hedging element of choice for 'native speakers of English' (Hyland, 1994:245; Varttala, 1999). In the context of this study then, a hedge can be thought of as 'a lexical device used by a writer to qualify a lack of commitment to the truth of a proposition he/she writes'.

The range and frequency of lexical hedging devices across the corpus was examined using a reference list of these items. Hyland and Milton (1997:187), using sources such as previous analysis of corpora, literature on modality, studies investigating expressions used to moderate the strength of arguments and claims, and reference grammars, compiled a list of 75 lexical hedges 'common to native speaker usage'. This list was subsequently expanded (Hyland, 2005) such that it numbered 101 items. This expanded list was chosen as the basis for the list of hedging items used in this study, since it was formulated from a more in-depth literature search than the 1997 list, was additionally informed by a corpus (K. Hyland, personal communication, Aug 3rd, 2019), and was also utilised by Vázquez & Giner (2008). However, in the present study, some changes were made. A number of items were added to the searches; this was either because their presence was suspected based on the presence of other, similar items on the list, or because they were identified through concordance searches for similar terms during analysis. A concordance search for 'assume*', for example, (* serving the purpose of a 'wildcard' representing any characters following the search term) would result in hits not only for *assume* and *assumed*, which appear on the original list, but also for *assumes*, which does not. Although different tenses of verb hedges were included in the searches, for the final count the different forms of a given verb were combined rather than treated as separate hedges. This is because the use, for example, of 'claim' represents the same choice of hedge as a use of 'claims' or 'claimed', with only the syntax of the sentence causing the variation. As this study focuses on the range and frequency of lexical hedges, rather than how they may be employed syntactically, counting different verb tenses as separate items would artificially reduce the overall count in terms of how frequently and widely a given verb hedge is used.

Additionally, two items from the original list – *would* and *wouldn't* - were removed. The status of *would* in terms of its function as a hedge is not uncomplicated. One of the most common uses of *would* in academic writing is as a hypothetical marker (Coates, 1983; Hyland, 1998), as in the following examples from Hyland (1998: 112), and this corpus, respectively:

'…if the haemoglobin in the root were mainly at the tip, it *would* have a local concentration 100-fold or greater than originally deduced'

'However, if they feel that they do not want the researcher to use their answers anymore after they have submitted the questionnaire, they *would* have to inform the researcher' (Spekafa1).

Uses such as those above suggest some premise or involve an expressed condition. As Varttala (2001) observes, while this can be considered as expressing conditional predictability and having a sense of epistemic meaning, this use does not convey tentativeness in the same way as other hedges. Rather than being epistemic 'in the sense of marking the information qualified by the auxiliary as uncertain…' here the tentativeness concerns 'whether or not the condition expressed by the if-clause is realized' (Varttala, 2001:111). Moreover, syntactically, *would* frequently co-occurs with verbs such as *appear*, *seem*, and *tend* (Huddleston, 1971; Hyland, 1998; Varttala, 2001), and arguably it is these verbs that constitute the primary hedge, rather than *would* itself. Even accepting that there may be cases where *would* can be considered a hedge in the terms defined in this study, previous research has shown that such usage occurs more frequently in spoken than written English, and may indeed be quite rare in academic writing (Butler, 1990; Coates, 1983; Varttala, 2001). It was for these reasons that the decision was made to exclude these two items from the analysis.

The result was a list of 133 lexical hedges (with different tenses of verbs combined following the searches, the final total was reduced to 91). This list, while not exhaustive, includes a wide variety of hedging devices, which were anticipated to span a considerable range in terms of frequency of use. Analysis of the items in the list will therefore serve to provide an informative picture of the frequency and range of lexical hedges employed within the corpus. Figure 1 details the list in full. Items added for this study are italicised.

About, almost, apparent, apparently, appear, appeared, appears, *appearance, appearing*, approximately, argue, argued, argues, *arguing*, around, assume, assumed, *assumedly, assumes*, broadly, certain amount, certain extent, certain level, *certain degree*, claim, claimed, claims, *claiming, contend, contended, contending, contends*, could, could not, doubt, *doubted, doubts*, doubtful, essentially, estimate, estimated, *estimates, estimating*, fairly, *feel*, feels, felt, *feeling*, frequently, from my perspective, from our perspective, from this perspective, generally, guess, indicate, indicated, indicates, *indicating, indication, indicative, indicator*, in general, *in many cases, in many instances*, in most cases, in most instances, in my opinion, *in our opinion*, in my view, in this view, in our view, largely, likely, mainly, may, maybe, might, mostly, often, on the whole, ought, perhaps, *plausible*, plausibly, possible, possibly, postulate, postulated, postulates, *postulating, presumable*, presumably, probable, probably, quite, rather x, relatively, roughly, *seem, seemed, seeming, seemingly*, seems, should, sometimes, somewhat, suggest, suggested, suggests, *suggestion, suggesting, suggestive*, suppose, supposed, *supposedly*, supposes, suspect, *suspected*, suspects, *suspecting*, tend to, tended to, tends to, *tending to*, to my knowledge, typical, typically, uncertain, uncertainly, unclear, unclearly, unlikely, usually.

Figure 1-List of hedging items used in this study (based on Hyland, 2005)

In each of the eight disciplines in the L1S, L2S and Published sub-corpora, these items were searched for using the Concord tool in Wordsmith Tools (Scott, 2019). This allowed the total number of individual usages to be quantified, and also displayed the terms in context, such that in cases where an item had potential uses other than as a hedge, these cases could be excluded. Regarding exclusions, these were not always easy determinations to make, and ideally a second 'rater' would have been used so as to seek agreement where examples were unclear. Unfortunately this was not possible, so I made all decisions on inclusion/exclusion, being as consistent as possible during the process. The guiding principle was that if, having

examined the context of use, a firm determination could not be made, the example was

excluded. Hedges that appeared within direct quotations from other reference sources were

also excluded, as these would not represent the original output of the writer themselves. There

has been discussion over whether hedges originating (but not directly quoted) from previous

work can be legitimately regarded as the output of the writer using them. Crompton (1997)

for example, suggests they cannot, since they do not represent the elected choice of the writer

if they originate within the proposition being reported on. However, I tend to favour Varttala's

(1999:185) view that, 'even when a tentative verb originates from the proposition reported,

the author in fact adopts the same tentative standpoint as the original source by choosing to

convey the same degree of tentativeness as the source'. In any case, as Varttala also observes,

determining the precise origin of all the hedges used to report on previous work would require

in-depth scrutiny of every book or research article referred to – a task that for obvious reasons

would be impracticable given the timeframe of this study. For this reason, hedges were

included unless found within direct quotations from previous sources. Once all hedges had

been quantified, the ten most commonly used in each sub-corpus were identified so as to allow

differences in usage across writer groups and disciplines to be assessed.

3.8.4 Citations

Concordance searches were carried out using the Concord utility in Wordsmith Tools (Scott,

2019), in order to identify in-text citations throughout the corpus. The L2S, L1S, and Published

sub-corpora in each discipline were examined individually. The following search terms were

used in an effort to identify the full variety of possible citation forms:

  (*) – Identifies integral citations, where year of publication is given in parentheses – e.g.,

  Author (2017).

(*,*) – Identifies non-integral citations containing author's name and year of publication – e.g., (Author, 2017).

* and *,*)/ * & *,*) – Identifies non-integral citations where multiple authors are attributed, followed by year of publication – e.g., (Author and Author, 2017) / (Author, Author and Author, 2017).

et al,*) – Identifies non-integral citations including et al – e.g., (Author et al, 2017).

;,*) – Identifies non-integral citations where multiple references are cited together – e.g., (Author, 2010; Author, 2011; Author, 2015).

Once concordance lists had been generated, these were checked in order to remove any irrelevant results, as the search terms also identified some cases in which the numbers or words in parentheses were not related to citations. The number of integral and non-integral citations in each sub-corpus was then established, so that these could be calculated as a percentage of total citations.

The next step in analysing citations in the corpus was to investigate structure and the use of verbs in integral citations. As referred to in Section 2.2.3.4, the taxonomies of Swales (1990) and Hyland (1999b, 2002b) are followed in this study. Swales' (1990) taxonomy classifies integral citations as either 'reporting', where the citation employs some kind of reporting verb:

'*Swales (1990) **notes** that this is a reporting citation'*

or as 'non-reporting', where some alternative syntactic structure is employed:

'*This would be a non-reporting citation in Swales' (1990) taxonomy'*

Hyland (1999b, 2002b) further sub-divided reporting citations by function. Activities described by reporting verbs can be classified as 'research acts' (e.g. - *analyse, prepare, interview*), 'cognition acts' (e.g. - *consider, view, conclude*) or 'discourse acts' (e.g. – *argue, suggest, note*).

Integral citations were therefore isolated, and all integral citations identified by the concordance searches were either recorded as non-reporting, or else the specific reporting verb used in the citation was recorded, such that for each sub-corpus the total number of different verbs employed in citations, as well as the number of times each individual verb was utilised could be ascertained. Each verb's use could then be calculated as a percentage of total integral citation use in the L2S, L1S and Published sub-corpora for each discipline, and in the entire corpus.

Having identified and quantified all of the reporting verbs associated with integral citations, the next step was to classify each verb according to Hyland's taxonomy. In most cases this was straightforward, although in some cases it was more problematic to discriminate between the research, cognitive, and discourse categorisations. The verb 'repeat' for example, could refer to the discourse act of restating the thoughts/arguments of a previous researcher, but could also refer to the process act of repeating an experimental procedure. It was therefore necessary in these cases, to return to the concordance results, identify the citations in which these problematic verbs had been used, and attempt to assess on an individual basis, which category a given verb use belonged to. In most instances this process facilitated clarification, and allowed verb uses to be categorised. A number of examples were encountered in which a given verb was found to fit into more than one category depending on the specific usage in question, and this was therefore an important process to conduct for the sake of accuracy in

112

categorisation. Despite successfully assigning most of the verb usages to categories, there remained cases in which a definitive decision could not be made, even given the additional context provided by the text. A fourth category was therefore created, such that these 'ambiguous' cases could be classified.

3.8.5 Semi-Structured Interviews

Once interviews were completed, they were transcribed and imported to the QSR NVivo qualitative data analysis software package. Here pertinent extracts could be identified, isolated, and grouped appropriately for later use.

3.8.6 Pre-sessional Materials

During examination of the materials used in the pre-sessional courses, references to academic vocabulary, lexical bundles, hedging, and citation were highlighted and noted so that they could be referred to following analysis of the results from the corpus. Particular attention was paid to any occurrence of the features as a primary focus of exercises/activities, although instances where reference was made as part of exercises/activities focused on other elements of academic writing were also noted.

4 Results

4.1 Overview of Chapter

The chapter is organised so as to answer the research questions as clearly as possible. Following top-level analyses of vocabulary, lexical bundles, hedging, and citation in the corpus as a whole, RQ1 (What is the extent of disciplinary variation in the use of academic vocabulary, lexical bundles, hedging and citations?) is addressed. Across-disciplines results for each of the four writing features are presented, focusing on the L2S, L1S and Published writer groups in turn. RQ2 (How do L2 student, L1 student, and published academic writing differ in the usage of academic vocabulary, lexical bundles, hedging and citations?) is addressed by comparisons of each writer group's usage of the four features in the corpus as a whole, followed by the results of within-disciplines analysis, wherein use of the writing features by each writer group within each of the eight disciplines is compared. Brief overviews of the analysis of existing pre-sessional materials and of the interview responses address RQ3 (How are academic vocabulary, lexical bundles, hedging and citations represented in existing pre-sessional writing materials?) and RQ4 (What are the perceptions of students who have completed the pre-sessional course in terms of the teaching of these four academic writing features, and their own difficulties with academic writing?), although a more detailed exploration of these two areas is provided in the discussion chapter).

4.2 The Corpus as a Whole

4.2.1 Vocabulary

As described in section 3.6.1, the use of vocabulary in the corpus was examined by means of the Range32 software package (Heatley and Nation, 2002). The software output made it

possible to determine which word families from the AWL occurred most frequently, and to determine the coverage provided by both the GSL and the AWL.

4.2.1.1 AWL and GSL Coverage

Table 6 below, shows the coverage provided by the first and second 1000 most frequent words from the GSL (combined to form a '1K and 2K' measure) and by the AWL, with respect to the corpus as a whole. The percentage of tokens and types in the corpus that do not appear on any of the three base lists is also indicated.

Table 6-GSL and AWL coverage of the whole corpus

Base List	Tokens %	Types %	Families
GSL 1K and 2K	75.67	11.98	998
AWL	11.93	5.15	570
Off-List	12.41	82.88	n/a

As the table shows, the AWL covers 11.93% of all the words in the corpus, and together with the GSL 1K and 2K lists, 87.6% of the words in the corpus are covered. What is not clear from an examination of the corpus as a whole however, is whether this coverage is uniform across disciplines or between writer groups within disciplines.

4.2.1.2 Frequency of AWL Vocabulary

Using the output from Range32, it was possible to ascertain the 20 most frequently occurring AWL families in the corpus. By far the most common family was *research* (includes *research, researcher, researchers, researched, researches,* and *researching*), with 13,088 occurrences, representing 2.84% of all AWL tokens in the corpus. Many of the 20 most frequently occurring AWL families may be considered general, research-related items (e.g., *data, process, approach,*

significant, and *theory*). Table 7, below, shows the 20 items in full, along with item % (calculated as family frequency divided by the total number of AWL tokens in the corpus – 461,320) and cumulative %. For the corpus as a whole, it can be seen that the 20 most frequently-occurring families account for 22.89% of all AWL tokens.

Table 7-20 most frequently occurring AWL families in the corpus as a whole

Family	Freq	item %	cum %
research	13088	2.84	2.84
data	8726	1.89	4.73
process	7166	1.55	6.28
positive	6046	1.31	7.59
approach	5743	1.24	8.84
individual	5509	1.19	10.03
significant	5211	1.13	11.16
focus	5022	1.09	12.25
media	4941	1.07	13.32
context	4621	1.00	14.32
specific	4507	0.98	15.30
role	4211	0.91	16.21
theory	4136	0.90	17.11
task	3999	0.87	17.98
range	3947	0.86	18.83
community	3862	0.84	19.67
impact	3831	0.83	20.50
previous	3726	0.81	21.31
target	3670	0.80	22.10
physical	3650	0.79	**22.89**

4.2.2 Lexical Bundles

The top 20 bundles in the corpus were, in many cases, suggestive of academic register, but were also subject nonspecific. Table 8, below, shows these bundles in full, along with total frequencies (F).

Table 8-Top 20 4-word bundles in the corpus as a whole

Bundle	F
it is important to	463
on the other hand	422
in the context of	360
as a result of	292
as well as the	290
at the same time	280
are more likely to	273
the end of the	258
at the end of	257
the extent to which	245
in the case of	226
one of the most	225
is one of the	218
to be able to	217
the purpose of this	198
in the form of	197
in relation to the	192
when it comes to	181
the use of the	173
in terms of the	168

4.2.3 Hedging

A total of 41,786 uses of the lexical hedges from the list detailed in Section 3.8.3 were identified in the whole corpus. Of the 91 different items counted from the list, 88 were present in the corpus. The top ten most commonly used hedges accounted for 66.11% of all hedge usage, and are shown in Table 9, along with frequencies (F) and percentage of total usage.

Table 9-Top 10 hedges in the corpus as a whole

Hedge	F	%
may	7083	16.95
suggest-	4189	10.02
could	3689	8.83
often	2303	5.51
likely	2221	5.32
might	2177	5.21
indicate-	2120	5.07
seem-	1432	3.43
argue-	1410	3.37
appear-	1000	2.39
TOTAL	27624	66.11

The top ten lexical hedges are dominated by lexical and modal verbs, with *may* by far the most common hedge used. Only two non-verb hedges make the top ten – the adverb *often*, and *likely*, which may be an adverb or adjective depending on usage.

4.2.4 Citation

The corpus as whole contained 31,361 citations, representing a total citation rate of 8.03 citations per 1000 words. The corpus contained 20,965 (66.28%) non-integral, and 10,666 (33.72%) integral citations, meaning that non-integral citations were almost twice as common.

Examining integral citations more closely, and following the taxonomies of Swales (1990) and Hyland (1999b, 2002b), a total of 2016 (representing 18.90%) of these were found to use a non-reporting structure (see Section 2.1.3.4 for details). The remainder utilised reporting verbs[2] in their structure. A total of 438 different reporting verbs were used in the corpus as a whole. Table 10 shows the 20 most commonly used reporting verbs in the corpus, with

---

[2] For the purposes of this analysis the preposition 'according to' is included under the label 'reporting verbs', since it serves a similar function and proved to be frequently used, amounting to 652 cases or 6.11% of all integral citations. It is not included in figures pertaining to discourse, cognition and research act verbs.

frequencies (F) and percentage of total integral citation use. These 20 verbs represent just over

50% of total integral citation use.

Table 10-20 most frequently used reporting structures in the corpus as a whole

Reporting Verbs	F	%
find	774	7.26
suggest	744	6.98
according to	652	6.11
state	486	4.56
argue	444	4.16
describe	278	2.61
note	217	2.03
report	200	1.88
show	186	1.74
propose	176	1.65
conduct	161	1.51
define	158	1.48
identify	149	1.41
claim	123	1.15
point out	110	1.03
explain	109	1.02
discuss	108	1.01
develop	102	0.96
support	97	0.91
use	94	0.88
Total	5368	50.34

Hyland's taxonomy classified reporting verbs as representing discourse acts (e.g., claim, argue, state) cognition acts (e.g., think, theorise, realise) or research acts (e.g., conduct, develop, demonstrate). As can be seen, discourse act verbs are common among the 20 most frequently used reporting structures. Indeed, throughout the corpus, discourse act verbs (totalling 4591) represented 43.04% of all integral citations. Cognition act verbs were far less frequent, totalling only 536 (5.03%), while research act verbs amounted to 2822 (26.46%). There were 49 ambiguous cases (0.46%), where reporting verbs in given citations could not be definitively categorised. These involved six verbs – introduce, reason, observe, equate, estimate, and show,

but represented few enough cases so as to have a negligible effect on overall percentages in each category.

4.3 RQ1 - Variation across Disciplines

4.3.1 Vocabulary

For the analysis of vocabulary use in and across the various sub-corpora, the decision was made to focus on the percentage coverage provided by each of the wordlists with specific reference to types. These were chosen over tokens as a focus since it was the variety of vocabulary employed by the writers that was of interest. For example, an individual utilising a given type from the AWL demonstrates sufficient knowledge of and familiarity with that type to be willing to employ it in their writing (the issue of whether vocabulary is *correctly* employed is not within the remit of this study). However, a greater knowledge and familiarity cannot necessarily be inferred if the same type is used twice, or indeed if it is used another 20 times, and thus for the purposes of this study, examining tokens would not have added additional value to the results.

There is however a major issue with examining types, as an initial inspection of the data highlighted - namely that the results obtained from Range32 for percentage AWL type coverage, GSL 1K and 2K type coverage, and Off-List type coverage, were text-length dependant. The effect of text length on various measures of lexis has been noted previously (Daller & Xue, 2007; Ryoo, 2018). It makes sense that text length should be influential on the measures in question in this case, given that there are a limited number of AWL and 1K/2K types (3107, 4119 and 3708 respectively). The percentage coverage of a text offered by the types in these lists is therefore likely, after a certain point, to decrease as a text gets longer. Conversely, given that Off-List types are not restricted in the same way, with increasing text

length, the percentage coverage of Off-List items is likely to increase. This can be clearly illustrated by presenting a selection of data from these initial analyses. Figure 2 below, shows the results from Range32 when L1S, L2S and Published writers were combined so as to produce an output for each discipline as a whole. The disciplines are presented in order of size (word count), increasing from left to right. There is a clear trend for percentage AWL types and percentage 1K and 2K types to decrease as text length increases, while percentage Off-List types follows the opposite pattern.
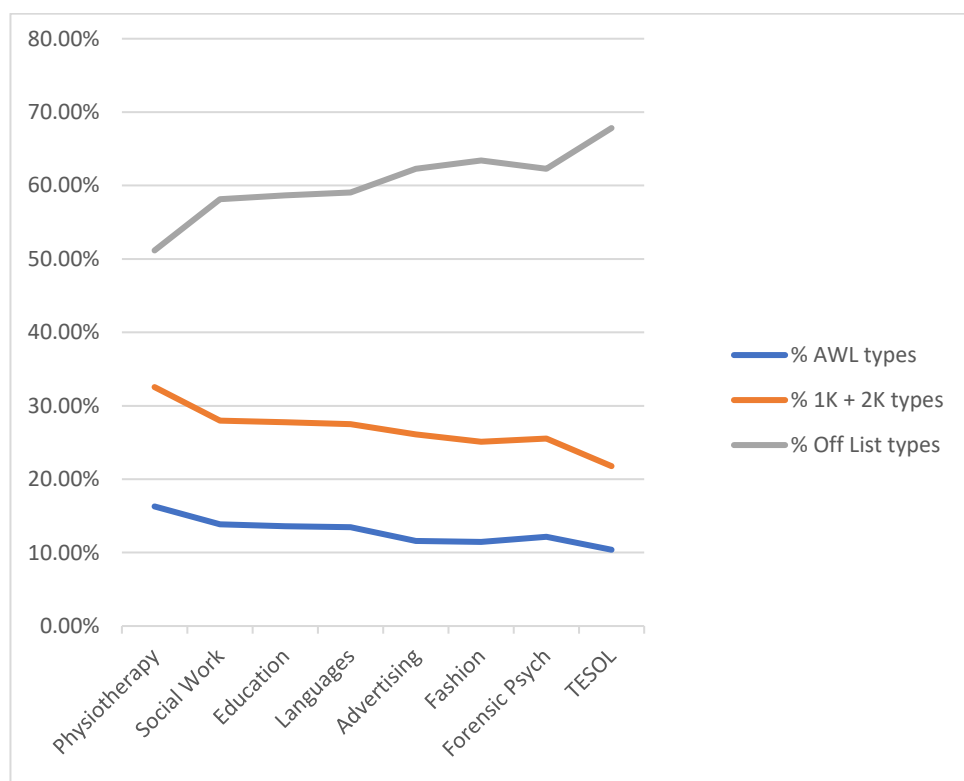
Figure 2-% AWL, 1K and 2K, and Off-List types by discipline

As a result of this text-length dependency, it was necessary to devise some method of normalisation before any valid comparisons of sub-corpora could be carried out, given that text length varies throughout the corpus, both in terms of disciplines and writer groups. One option would be to take the smallest sub-corpus in any given comparison, and then extract samples of text of approximately equal size from the other sub-corpora to facilitate the

comparison. For example, among L1 writers, the smallest sub-corpus was that of Physiotherapy, which totalled 22,492 words. In order to compare L1 writers in each discipline, samples of approximately 22,492 words could be taken at random from the L1 sub-corpora in each of the other seven disciplines. This would create eight equally sized texts, which could then be compared using Range32, without the problem of word count variation influencing the output. However, while this method would indeed have removed the problem of text length, it would also have created issues of representativeness and data loss. It would be unwise to assume for example, that a 22,492-word extract from the 338,100-word L1 TESOL sub-corpus would be representative of that sub-corpus as a whole, and the failure to include over 300,000 words of text would constitute a substantial loss of potential data. A more suitable alternative was therefore required.

It was clearly still necessary to take the smallest sub-corpus in any given comparison as a baseline from which to normalise the other sub-corpora to be compared. However, in order to improve representativeness and avoid data loss, rather than taking a single comparable extract from each of the larger sub-corpora, a series of extracts (as many as the word count would allow) of approximately equal length to the smallest sub-corpus in the comparison was taken from each. Each of these extracts was then analysed separately in Range32, and for a given sub-corpus the mean values of % AWL types, % 1K and 2K types, % Off-List types, and AWL families from all of these analyses were calculated, thus giving values for an 'average extract'. For example, returning to the comparison of L1 writers in each discipline, the 22,492-word Physiotherapy sub-corpus represents the smallest in this comparison. In order to meaningfully compare this to the 140,622-word L1 Advertising sub-corpus, six extracts of ~22,492 (+/- up to 20 words) were taken from the larger sub-corpus, and after each was analysed in Range32,

values for the 'average 22,492' words were calculated. This procedure of identifying the smallest sub-corpus in any given comparison and then producing results for comparable average extracts from the larger sub-corpora was repeated throughout the vocabulary analysis.

4.3.1.1 Vocabulary Use by L2S Writers across Disciplines

For the AWL/GSL type analysis, the L2S sub-corpora for each discipline were normalised as described in Section 4.3.1, here being balanced down to match the smallest sub-corpus – Physiotherapy at 6590 words.

4.3.1.1(i) AWL and GSL Coverage

Table 11 below, shows the coverage provided by the GSL 1K and 2K lists and by the AWL, with respect to the eight L2S disciplines. Type percentages from each base list are presented, along with Off-List type percentages and total AWL families.

Table 11-AWL and GSL coverage for L2S writers in each discipline

	L2S Writers by Discipline							
	Advertising	Education	Fashion	Forensic Psychology	Languages	Physiotherapy	Social Work	TESOL
GSL 1K and 2K type %	59.29	60.52	57.66	57.15	63.64	57.41	56.22	58.37
AWL type %	18.17	20.73	17.46	21.18	18.75	22.75	20.77	20.62
Off-List type %	22.55	18.75	24.89	21.67	17.62	19.84	23.02	21.01
AWL Families	199	190	190	203	180	220	197	189

Among L2S writers GSL 1K and 2K coverage varied across the eight disciplines, with a 7.42% difference between the lowest coverage in Social Work and the highest in Languages. AWL coverage exhibited less variation, with a range of only 5.29% separating the minimum in Fashion and maximum in Physiotherapy. As well as showing the minimum AWL coverage,

123

Fashion also exhibited the maximum Off-List coverage, while minimum Off-List coverage was found in Languages, the discipline that also demonstrated the highest 1K and 2K coverage. It is important to note however, that Off-List items may represent a variety of things – some may be technical or subject-specific vocabulary, but since it was not possible given the size of the corpus and the time restraints of the study to 'clean up' the off-list items, some may also represent proper nouns, non-English words, or spelling errors. In terms of AWL families, although AWL type coverage was lowest in Fashion, it was Languages that utilised the fewest AWL families overall, with the maximum found in Physiotherapy. It should be noted that the reason the overall number of AWL families is substantially lower than in the preceding analyses, is simply due to the smaller size of the 'average' extracts of the sub-corpora used in this analysis.

4.3.1.1(ii) Frequency of AWL Vocabulary

To examine the use of AWL families, the L2S sub-corpora across the eight disciplines were analysed in their entirety. Table 12 shows the top 20 AWL families for L2S writers in each discipline, with frequencies and cumulative percentages, and indicates the extent of item sharing across disciplines. Yellow shading indicates items shared between 5-7 disciplines, blue shading shows those families shared between 2-4 disciplines. Other items were unique to one discipline.

Table 12-20 most frequently occurring AWL families for L2S writers by discipline

Advertising			Education			Fashion			Forensic Psychology		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	159	3.31	research	1010	10.40	range	229	4.68	research	213	4.62
create	136	6.14	method	208	12.54	design	210	8.98	individual	118	7.18
consume	103	8.29	participate	199	14.58	consume	193	12.92	identify	98	9.30
philosophy	99	10.35	data	181	16.45	strategy	144	15.87	factor	95	11.36
culture	98	12.39	individual	139	17.88	create	113	18.18	media	86	13.23
ethic	76	13.97	focus	129	19.21	sustain	97	20.16	analyse	73	14.81
identify	73	15.49	approach	128	20.52	finance	74	21.68	psychology	65	16.22
image	72	16.99	process	125	21.81	analyse	73	23.17	evident	63	17.59
focus	71	18.47	institute	125	23.10	label	70	24.60	mental	58	18.85
gender	69	19.90	policy	124	24.37	style	67	25.97	data	57	20.08
individual	67	21.30	communicate	122	25.63	ratio	66	27.32	domestic	57	21.32
media	63	22.61	implement	120	26.86	margin	64	28.63	community	56	22.53
minor	58	23.82	achieve	114	28.04	trend	64	29.94	technique	56	23.75
ethnic	56	24.98	ethic	112	29.19	culture	56	31.08	culture	51	24.85
target	54	26.11	analyse	109	30.31	append	53	32.17	investigate	51	25.96
communicate	51	27.17	require	97	31.31	item	49	33.17	positive	50	27.04
role	50	28.21	issue	96	32.30	process	48	34.15	approach	50	28.13
impact	48	29.21	gender	87	33.19	tradition	45	35.07	theme	47	29.15
specific	46	30.17	topic	86	34.08	media	43	35.95	perceive	47	30.17
perceive	46	**31.13**	define	85	**34.95**	target	42	**36.81**	respond	46	**31.16**

Languages			Physiotherapy			Social Work			TESOL		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
strategy	208	2.52	team	67	8.38	theory	176	7.68	participate	886	3.11
target	191	4.84	culture	33	12.50	focus	75	10.95	research	563	5.08
task	186	7.10	role	26	15.75	approach	62	13.65	append	518	6.89
append	154	8.97	task	18	18.00	individual	53	15.96	analyse	517	8.71
focus	144	10.72	positive	17	20.13	intervene	45	17.92	motive	512	10.50
create	138	12.39	distribute	16	22.13	issue	43	19.80	data	495	12.24
instruct	133	14.01	implement	15	24.00	assess	42	21.63	communicate	489	13.95
environment	119	15.45	individual	14	25.75	adult	38	23.29	identify	438	15.48
achieve	119	16.90	challenge	13	27.38	identify	36	24.86	method	429	16.99
motive	118	18.33	structure	12	28.88	furthermore	35	26.38	strategy	418	18.45
challenge	111	19.67	achieve	11	30.25	psychology	34	27.87	task	400	19.86
communicate	108	20.99	establish	11	31.63	professional	30	29.18	culture	400	21.26
positive	104	22.25	analyse	10	32.88	attach	29	30.44	attitude	389	22.62
factor	103	23.50	create	10	34.13	impact	28	31.66	context	370	23.92
process	101	24.72	vision	9	35.25	process	23	32.66	process	310	25.00
culture	93	25.85	goal	9	36.38	aspect	23	33.67	approach	306	26.08
research	92	26.97	involve	9	37.50	occur	23	34.67	evaluate	290	27.09
approach	90	28.06	outcome	9	38.63	partner	21	35.59	focus	268	28.03
individual	81	29.04	professional	8	39.63	analyse	21	36.50	assess	245	28.89
topic	80	**30.02**	strategy	8	**40.63**	environment	20	**37.37**	status	122	**29.32**

No top 20 AWL families were shared by all eight or even by seven disciplines, and only three families - *culture, individual* and *analyse* were shared by six disciplines. Research was the most commonly used family in three disciplines – Advertising, Education, and Forensic Psychology. Families shared between 5-7 disciplines included items related to research in general – *research, analyse, approach, focus*, and *process* for example. This description also fits a number of the items shared between 2-4 disciplines, such as *impact, positive, identify, method* and *data*.

Items unique to only one discipline included: Advertising - *philosophy, image, ethnic*; Education - *institute, policy*; Fashion - *range, design, label, trend*; Forensic Psychology - *evident, mental, domestic, community*; Languages - *instruct, challenge*; Physiotherapy - *team, goal, outcome, structure*; Social Work - *intervene, adult, attach, partner*; and TESOL - *attitude, context, evaluate.* This suggests that there is a degree of subject specificity in the most commonly used AWL families in the disciplines.

Table 13 shows the extent of item sharing between disciplines for L2S writers, as percentages of the total usage within the top 20 items (i.e., the sum of all frequencies within the top 20 is regarded as representing 100%, and therefore in a given discipline, sharing between 5-7 disciplines for example, is arrived at by adding all the frequencies of AWL families falling into that category, and calculating what percentage of all the usage within the top 20 that total represents).

Table 13-Top 20 AWL families item sharing for L2S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 20 usage	1495/100	3396/100	1800/100	1437/100	2473/100	325/100	857/100	8365/100
Shared between all disciplines frequency / % of top 20 usage	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Shared between 5-7 disciplines frequency / % of top 20 usage	395/26.42	1640/48.29	177/9.83	505/35.14	601/24.30	57/17.54	234/27.30	2364/28.26
Shared between 2-4 disciplines frequency / % of top 20 usage	769/51.44	1325/39.02	588/32.67	498/34.66	1628/65.83	113/34.77	233/27.19	4830/57.74
Total shared % of top 20 usage	71.86	87.31	42.50	69.80	90.13	52.31	54.49	86.00
Unshared frequency / % of top 20 usage	331/22.14	431/12.69	1035/57.50	434/30.20	244/9.87	155/47.69	390/45.51	1171/14.00

The disciplines in which the largest proportion of top 20 usage was made up of items shared between 5-7 disciplines were Education (48.29%) and Forensic Psychology (35.14%), suggesting that, compared to other disciplines, common AWL usage in these subjects may tend more towards those families applicable to a variety of disciplines. Physiotherapy (47.69%), Social Work (45.51%) and Fashion (57.50) all made up the highest proportion of their top 20 usage from items unique to the discipline, and there is some evidence of potential subject specificity among these items. Overall sharing was highest in Languages, in which 90.13% of

top 20 usage represented shared items. Lowest sharing overall was found in Fashion, at 42.50%. These results would seem to suggest that the degree to which the academic vocabulary most commonly used by L2S writers can be considered 'generic' differs across disciplines.

4.3.1.2 Vocabulary Use by L1S Writers across Disciplines

For the AWL/GSL type analysis, the L1S sub-corpora for each discipline were normalised as described in Section 4.3.1, here being balanced down to match the smallest sub-corpus – Physiotherapy at 22,492 words.

4.3.1.2(i) AWL and GSL Coverage

The coverage provided by the GSL 1K and 2K lists and by the AWL, with respect to the eight L1S disciplines is shown in Table 14, below. Type percentages from each base list are presented, along with Off-List type percentages and total AWL families.

Table 14-AWL and GSL coverage for L1S writers in each discipline

	L1S Writers by Discipline							
	Advertising	Education	Fashion	Forensic Psychology	Languages	Physiotherapy	Social Work	TESOL
GSL 1K and 2K type %	51.42	56.90	50.17	49.61	53.82	47.24	51.78	50.15
AWL type %	17.17	19.35	17.66	20.27	19.90	20.86	21.16	21.37
Off-List type %	31.41	23.76	32.18	30.11	26.28	31.90	27.06	28.48
AWL Families	311	306	325	323	337	352	312	359

GSL 1K and 2K coverage was highest in Education and lowest in Physiotherapy. Coverage provided by AWL types ranged from 21.37% in TESOL, down to just over 17% in Advertising. AWL families were also highest in TESOL, but were lowest in Education, which also showed the

128

lowest Off-List coverage. As in the L2S analysis, Fashion was the discipline with the highest Off-List type coverage.

4.3.1.2(ii) Frequency of AWL Vocabulary

The L1S sub-corpora across the eight disciplines were examined in their entirety for this analysis. The top 20 AWL families for L1S writers in each discipline are shown in Table 15, along with frequencies and cumulative percentages. Yellow shading indicates items shared between 5-7 disciplines, blue shading shows those families shared between 2-4 disciplines. Other items were unique to one discipline.

Table 15-20 most frequently occurring AWL families for L1S writers by discipline

Advertising			Education			Fashion			Forensic Psychology		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	543	4.12	research	370	9.19	range	730	4.98	research	1908	5.78
create	374	6.96	participate	105	11.80	consume	603	9.10	identify	958	8.69
gender	325	9.43	method	78	13.74	create	315	11.25	theme	896	11.40
target	277	11.53	achieve	73	15.55	strategy	282	13.17	individual	895	14.11
analyse	268	13.57	focus	68	17.24	design	272	15.03	analyse	758	16.41
consume	268	15.60	approach	63	18.81	trend	245	16.70	evident	681	18.48
media	267	17.63	create	61	20.32	finance	220	18.20	media	666	20.49
ethic	219	19.29	data	59	21.79	style	219	19.69	data	618	22.37
identify	213	20.91	involve	55	23.16	technology	198	21.05	factor	580	24.13
image	200	22.43	access	54	24.50	focus	174	22.23	mental	441	25.46
culture	192	23.88	ensure	54	25.84	sustain	172	23.41	sex	435	26.78
generation	181	25.26	individual	54	27.18	analyse	167	24.55	participate	354	27.85
focus	170	26.55	benefit	48	28.37	target	166	25.68	previous	326	28.84
role	163	27.79	environment	47	29.54	margin	153	26.72	partner	315	29.80
communicate	139	28.84	issue	44	30.63	media	148	27.73	negate	315	30.75
similar	134	29.86	process	41	31.65	invest	147	28.74	issue	291	31.63
feature	131	30.85	specific	40	32.65	item	142	29.71	physical	280	32.48
previous	126	31.81	paradigm	39	33.61	ratio	136	30.63	proceed	275	33.32
range	118	32.71	require	38	34.56	purchase	136	31.56	conduct	272	34.14
brief	116	**33.59**	ethic	36	**35.45**	indicate	125	**32.41**	psychology	264	**34.94**

Languages			Physiotherapy			Social Work			TESOL		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
target	189	2.61	team	95	3.64	theory	798	9.21	participate	958	2.48
task	178	5.06	culture	73	6.44	approach	341	13.15	research	927	4.87
append	169	7.39	individual	66	8.98	individual	292	16.52	analyse	764	6.84
focus	150	9.46	evident	66	11.51	focus	210	18.95	culture	756	8.80
strategy	136	11.34	injure	53	13.54	intervene	198	21.24	data	675	10.54
challenge	119	12.98	role	51	15.50	issue	175	23.26	motive	601	12.09
differentiate	109	14.48	approach	41	17.07	impact	129	24.75	identify	538	13.48
create	102	15.89	professional	39	18.57	attach	127	26.21	append	486	14.74
achieve	100	17.27	challenge	36	19.95	domestic	118	27.58	assess	471	15.96
communicate	100	18.65	style	35	21.29	method	118	28.94	focus	453	17.13
ensure	94	19.94	focus	35	22.63	identify	117	30.29	task	446	18.28
environment	88	21.16	outcome	35	23.97	positive	110	31.56	method	433	19.40
motive	88	22.37	significant	32	25.20	role	104	32.76	respond	430	20.51
approach	85	23.54	assess	32	26.43	recover	97	33.88	communicate	423	21.60
instruct	81	24.66	promote	30	27.58	mental	94	34.97	strategy	396	22.63
individual	72	25.66	identify	30	28.73	adult	91	36.02	attitude	395	23.65
require	72	26.65	demonstrate	28	29.80	professional	85	37.00	specific	392	24.66
academy	72	27.64	research	26	30.80	demonstrate	79	37.91	approach	383	25.65
factor	71	28.62	potential	26	31.80	perspective	77	38.80	context	371	26.61
assess	69	**29.57**	strategy	26	**32.80**	assess	73	**39.64**	individual	355	**27.52**

As with the L2S analysis, no AWL families appeared in the top 20 for all eight disciplines. There are also other notable similarities. In all disciplines except Languages, the most frequently occurring family for L2S writers was the same as that for L1S writers. Languages and TESOL had the highest number of items shared overall, with only three families unique to each discipline. At the other end of the scale, Fashion showed the highest number of unique items – 11 out of 20. In all disciplines, sharing between 2-4 other disciplines was more common (in terms of number of items) than sharing between 5-7 disciplines.

Again, items commonly shared centred around general research terminology – *research, focus, process, approach* and *individual* being examples. AWL families unique to a single discipline included examples that may be somewhat more discipline related, such as *gender, brief* and *image* (Advertising); *design, purchase* and *trend* (Fashion); *domestic* and *intervene* (Social Work); and *team, outcome* and *injure* (Physiotherapy); however, there were varying degrees of consistency between L2S and L1S writers in terms of the specific families unique to each discipline. Fashion showed the most commonality, with *design, sustain, finance, ratio, margin* and *trend* occurring among unique items for both writer groups. L2S and L1S writers in Social Work shared four unique items - *theory, intervene, adult* and *attach,* while in TESOL, the only common families were *attitude* and *context.* Four disciplines - Advertising, Forensic Psychology, Languages, and Physiotherapy showed only one family common to the unique items for both writer groups (*image, theme, instruct* and *team,* respectively). No unique-to-discipline items were found to be common to both L2S and L1S writers in Education.

The extent of item sharing between disciplines for L1S writers, expressed as percentages of the total usage within the top 20 items, can be seen in Table 16. In Social Work the highest proportion of top 20 usage came from unique-to-discipline items, whereas in all other

131

disciplines items shared between 2-4 disciplines made up the largest proportion of top 20 usage, suggesting that while common AWL use by L1S writers in these disciplines may not tend towards strict subject specificity, it is also not predominantly made up of the broadest, generally applicable families. Highest overall item sharing was found in TESOL (88.77%) and Languages (87.78%), with Social Work showing the lowest proportion of item sharing overall (49.17%), followed by Fashion, at 59.03%.

Table 16-Top 20 AWL families item sharing for L1S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 20 usage	4424/100	1427/100	4750/100	11528/100	2144/100	855/100	3433/100	10653/100
Shared between all disciplines frequency / % of top 20 usage	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Shared between 5-7 disciplines frequency / % of top 20 usage	926/20.93	555/38.89	174/3.66	3761/32.62	307/14.32	198/23.16	960/27.96	2656/24.93
Shared between 2-4 disciplines frequency / % of top 20 usage	2411/54.50	635/44.50	2630/55.37	4715/40.90	1575/73.46	386/45.15	728/21.21	6801/63.84
Total shared % of top 20 usage	75.43	83.39	59.03	73.52	87.78	68.31	49.17	88.77
Unshared frequency / % of top 20 usage	1087/24.57	237/16.61	1946/40.97	3052/26.47	262/12.22	271/31.70	17.45/50.83	1196/11.23

4.3.1.3 Vocabulary Use by Published Writers across Disciplines

The Published sub-corpora for each discipline were normalised as described in Section 4.3.1

for the AWL/GSL type analysis, here being balanced down to match the smallest sub-corpus –

Physiotherapy at 167,298 words.

4.3.1.3(i) AWL and GSL Coverage

The data in Table 17 show the percentage coverage of GSL 1K and 2K, and AWL types, for

writers in the Published sub-corpora across the eight disciplines. Off-list types and AWL

families are also shown.

Table 17-AWL and GSL coverage for Published writers in each discipline

	Published Writers by Discipline							
	Advertising	Education	Fashion	Forensic Psychology	Languages	Physiotherapy	Social Work	TESOL
GSL 1K and 2K type %	32.92	33.21	34.26	32.69	33.86	33.70	32.66	33.00
AWL type %	16.25	17.42	15.70	16.08	18.02	16.79	16.11	16.20
Off-List type %	50.83	49.37	50.04	51.24	48.12	49.31	51.23	50.81
AWL Families	530	540	541	540	529	503	541	528

Among Published writers, the highest 1K and 2K coverage was found in Fashion, which also

showed the lowest AWL type coverage. Perhaps surprisingly in light of this, the highest number

of AWL families was found in Fashion and Social Work. Off-List type coverage was highest in

Forensic Psychology and lowest in Languages – the discipline that also showed the highest AWL

type coverage.

4.3.1.3(ii) Frequency of AWL Vocabulary

The Published sub-corpora across the eight disciplines were examined in their entirety for this analysis. The top 20 AWL families for Published writers in each discipline are shown in Table 18 (along with frequencies, cumulative percentages, and item sharing. In the table, a white background indicates items shared between all eight disciplines. Other shading is as per previous tables. Among Published writers, four top 20 items (*research, analyse, data,* and *respond*) were found to be shared between all eight disciplines – a commonality not seen in either L2S or L1S writing. Sharing of items between 5-7 disciplines was also slightly more common in terms of the number of items; with the minimum in this category being four items shared (compared to three and one for L2S and L1S writers respectively). Similarly to the other writer groups, shared items tended to be general, research-related families, while items that were not shared could be seen in some cases to be rather more discipline related. Examples of this include *mental* and *incidence* in Forensic Psychology; *instruct*, *interact* and *motive* in TESOL; and *physical, injure* and *medical* in Physiotherapy. Of all the disciplines, Physiotherapy shared the fewest items, while Social Work shared the most – indeed every item in the top 20 was shared in this discipline.

  In terms of whether the writer groups showed similarity in the items that were unique to one discipline, there was very little commonality between Published and L2S writers, with only *mental* in Forensic Psychology and *team* in Physiotherapy being common to both groups. When Published writers are compared to L1S writers, the overlap was greater, with *academy* in Languages, and *injure*, *outcome, and team* in Physiotherapy, but the level of commonality between Published and the other writer groups in unique-to-discipline items was generally low nonetheless.

Table 18-20 most frequently occurring AWL families for Published writers by discipline

Advertising			Education			Fashion			Forensic Psychology		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
consume	2138	5.05	research	865	2.74	consume	1394	3.45	participate	912	3.33
research	1159	7.79	participate	743	5.10	research	925	5.75	research	639	5.67
perceive	922	9.97	data	563	6.88	analyse	622	7.29	significant	565	7.73
respond	882	12.06	analyse	539	8.59	strategy	540	8.63	individual	451	9.38
media	840	14.04	focus	488	10.14	perceive	510	9.89	analyse	448	11.01
purchase	781	15.89	culture	424	11.48	process	465	11.04	factor	377	12.39
attitude	773	17.72	identify	400	12.75	invest	461	12.18	approach	365	13.73
positive	743	19.47	significant	346	13.85	positive	452	13.30	respond	362	15.05
individual	731	21.20	context	324	14.88	vary	451	14.42	mental	361	16.37
significant	671	22.79	design	324	15.90	focus	446	15.53	vary	355	17.66
identify	605	24.22	individual	311	16.89	respond	446	16.63	identify	348	18.93
vary	579	25.58	approach	309	17.87	corporate	432	17.70	incidence	318	20.10
process	536	26.85	process	300	18.82	institute	422	18.75	assess	315	21.25
analyse	535	28.12	method	295	19.75	create	421	19.79	data	299	22.34
create	489	29.27	perceive	294	20.69	innovate	416	20.82	strategy	276	23.35
participate	478	30.40	community	289	21.60	identify	409	21.84	consist	270	24.33
communicate	467	31.51	project	288	22.52	theory	401	22.83	indicate	261	25.29
strategy	445	32.56	respond	274	23.38	resource	401	23.82	sex	257	26.23
data	417	33.54	involve	263	24.22	significant	385	24.78	specific	256	27.16
environment	410	**34.51**	vary	260	**25.04**	data	381	**25.72**	process	253	**28.09**

Languages			Physiotherapy			Social Work			TESOL		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	1075	3.13	participate	866	4.06	*research*	1100	2.54	task	1106	2.52
participate	856	5.63	injure	744	7.55	participate	790	4.37	participate	1029	4.86
process	617	7.43	physical	674	10.72	community	561	5.66	research	1006	7.16
assess	580	9.12	data	450	12.83	culture	543	6.92	analyse	809	9.00
policy	544	10.70	intervene	417	14.79	respond	543	8.17	context	679	10.55
context	529	12.24	analyse	413	16.72	identify	432	9.17	respond	655	12.04
analyse	472	13.62	assess	411	18.65	theory	420	10.14	instruct	617	13.44
approach	460	14.96	factor	368	20.38	individual	419	11.11	focus	571	14.74
focus	452	16.28	significant	324	21.90	positive	403	12.04	interact	558	16.02
professional	429	17.53	outcome	317	23.39	analyse	379	12.92	data	555	17.28
data	414	18.73	vary	303	24.81	significant	376	13.79	text	532	18.49
task	371	19.81	medical	277	26.11	focus	363	14.63	significant	499	19.63
identify	357	20.85	function	275	27.40	sex	362	15.46	identify	487	20.74
evaluate	324	21.80	research	233	28.49	process	345	16.26	process	481	21.83
text	318	22.72	specific	217	29.51	data	338	17.04	communicate	470	22.91
specific	311	23.63	evaluate	212	30.50	vary	325	17.79	vary	434	23.89
academy	307	24.53	team	198	31.43	professional	321	18.53	construct	376	24.75
individual	303	25.41	individual	196	32.35	intervene	313	19.26	item	368	25.59
respond	301	26.29	respond	195	33.27	factor	304	19.96	create	358	26.41
strategy	299	**27.16**	previous	189	**34.15**	involve	304	**20.66**	motive	337	**27.17**

Item sharing for Published writers in terms of percentage of total top 20 usage is shown in Table 19.

Table 19-Top 20 AWL families item sharing for Published writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 20 usage	14601/100	7899/100	10380/100	7688/100	9319/100	7279/100	8941/100	11927/100
Shared between all disciplines frequency / % of top 20 usage	2993/20.50	2241/28.37	2374/22.87	1748/22.74	2262/24.27	1291/17.74	2360/26.40	3025/25.36
Shared between 5-7 disciplines frequency / % of top 20 usage	3600/24.66	2848/36.06	2156/20.77	2884/37.51	2585/27.74	1689/23.20	3050/34.11	3501/29.35
Shared between 2-4 disciplines frequency / % of top 20 usage	5204/35.64	1903/24.09	3718/35.82	1846/24.01	3621/38.86	3621/22.32	3531/39.49	3145/26.37
Total shared % of top 20 usage	80.80	88.52	79.46	84.26	90.87	63.26	100	81.08
Unshared frequency / % of top 20 usage	2804/19.20	907/11.48	2132/20.54	1210/15.74	851/9.13	2674/36.74	0/0.00	2256/18.92

Four disciplines (Advertising, Fashion, Languages and Social Work) showed their highest proportion of top 20 usage within the 2-4 category, which may indicate that in these areas, AWL use by Published writers predominantly involves families that are neither subject specific nor applicable to a wide range of disciplines. Education, Forensic Psychology, and TESOL showed higher usage in the 5-7 category, suggesting a greater degree of subject

generalisability to their AWL use. Only Physiotherapy had unshared items as the highest percentage (36.74%), and a number of these items did demonstrate subject specificity. This may be a result of the rather more 'scientific' nature of this discipline compared to the others. Item sharing overall was highest in Social Work (100%).

4.3.1.4 Summary Comparison of Writer Groups across Disciplines

4.3.1.4(i) AWL/GSL Coverage

In terms of comparing the patterns of high and low coverage between the three writer groups, it can be seen that there are notable differences. Table 20, below, summarises the findings by indicating which disciplines in each writer group had the highest and lowest coverage in the various categories. Areas where writer groups share commonality are shaded.

Table 20-Highest and lowest GSL/AWL coverage across writer groups

	GSL 1K and 2K type %		AWL type %		Off-List type %		AWL Families	
	Highest	**Lowest**	**Highest**	**Lowest**	**Highest**	**Lowest**	**Highest**	**Lowest**
L2S	Languages	Social Work	Physiotherapy	Fashion	Fashion	Languages	Physiotherapy	Languages
L1S	Education	Physiotherapy	TESOL	Advertising	Fashion	Education	TESOL	Education
Published	Fashion	Social Work	Languages	Fashion	Forensic Psy	Languages	Social Work/Fashion	Physiotherapy

The three writer groups showed different patterns of type coverage, although the most similarities were between L2S and Published writers. In both of these writer groups the discipline with the lowest 1K and 2K coverage was Social Work, the lowest AWL coverage was found in Fashion, and the lowest Off-List coverage occurred in Languages. By contrast, L1S writers showed no commonality with Published writers, and only one point of similarity with L2S writers, with both of these writer groups showing the highest Off-List coverage in Fashion.

137

4.3.1.4(ii) Item Sharing

Table 21 shows the figures for total shared % usage for all three writer groups.

Table 21-Total sharing (% of total top 20 usage) for L2S, L1S and published writers

	Writer Group	Discipline							
		Advertising	Education	Fashion	Forensic Psy	Languages	Physiotherapy	Social Work	TESOL
Total Shared %	L2S	71.86	87.31	42.50	69.80	90.13	52.31	54.49	86.00
	L1S	75.43	83.39	59.03	73.52	87.78	68.31	49.17	88.77
	Published	80.80	88.52	79.46	84.26	90.87	63.26	100	81.08

As the table illustrates, there was no consistency between writer groups in terms of which disciplines had the highest or lowest proportions of shared items within the top 20. For L2S the disciplines showing the highest and lowest sharing proportions were Languages (90.13%) and Fashion (42.50%); for L1S writers these were TESOL (88.77%) and Social Work (49.17%); and for Published writers Social Work showed the highest proportion of sharing (100%), with Physiotherapy lowest (63.26). In terms of RQ1, these results show, where the most frequently used AWL vocabulary is concerned, that while some disciplines show high levels of commonality, for others, the extent of disciplinary variation is much greater. The disciplines were also not consistent in terms of which writer groups showed higher or lower levels of sharing. In Advertising, Fashion, and Forensic Psychology the proportions of sharing followed a L2S<L1S<Published pattern. In Education, Languages and Social Work a pattern of L1S<L2S<Published was found. Physiotherapy showed a L1S<Pub<L2S pattern, and finally usage within the TESOL discipline displayed a Pub<L2S<L1S pattern. This lack of consistency across both disciplines and writers groups would tend to suggest that the use of common AWL families cannot be considered uniform, either in terms of disciplines or writers.

138

4.3.2 Lexical Bundles

A total of 3916 bundles were used by L2S writers, giving an across disciplines average of 5.62 bundles per 1000 words. L1S writers' rate of use was slightly higher, with 8077 total bundles and 7.97 bundles per 1000 words. Of the three groups, overall bundle use was proportionally lowest in Published writers, with 6716 total bundles, and a rate of 3.20 bundles per 1000 words. Table 22 shows data for bundle use in each discipline and writer group. Bundle types and tokens are shown, along with figures for percentage of total words occurring in bundles, and bundles per 1000 words.

Table 22-Overall lexical bundle use in sub-corpora

		Advertising	Education	Fashion	Forensic Psych	Languages	Physio	Social Work	TESOL
L2S	Word Count	49696	88678	53246	38812	102370	6590	19033	262512
	Bundle Types	46	136	65	51	123	0	16	62
	Bundle Tokens	196	948	323	221	1045	0	79	1104
	% of words in bundles	1.58	4.28	2.43	2.28	4.08	0.00	1.66	1.68
	Bundles/1000 words	3.94	10.69	6.07	5.69	10.21	0.00	4.15	4.21
L1S	Word Count	140622	49870	150677	276292	82230	22492	80168	338100
	Bundle Types	77	82	68	138	146	16	242	54
	Bundle Tokens	655	406	684	2640	997	62	1491	1142
	% of words in bundles	1.86	3.26	1.82	3.82	4.85	1.10	7.44	1.35
	Bundles/1000 words	4.66	8.14	4.54	9.56	12.12	2.76	18.60	3.38
Pub	Word Count	302083	265841	303978	218408	287310	167298	267995	366970
	Bundle Types	45	58	39	50	47	77	51	34
	Bundle Tokens	882	947	776	716	898	770	861	866
	% of words in bundles	1.17	1.42	1.02	1.31	1.25	1.84	1.29	0.94
	Bundles/1000 words	2.92	3.56	2.55	3.28	3.13	4.60	3.21	2.36

L2S writers showed notable higher rates of bundle usage in Education and Languages compared to other disciplines, while Languages and, in particular, Social Work showed the highest rates among the L1S group. Rates among Published writers were much more uniform across disciplines. In term of the writer groups themselves within disciplines, Published writers showed the lowest rates of bundle use in every discipline in terms of the percentage of words

in bundles, and the lowest rate in seven of the eight disciplines in terms of bundles per 1000 words. Where the two student groups were concerned, the disciplines varied, with L2S writers showing higher rates in Education, Fashion and TESOL, and L1S rates being higher in the remainder of disciplines (with the exception of Physiotherapy where no L2S bundles were identified).

In the following sections, the results of the bundle analyses for individual writer groups across disciplines will be examined.

4.3.2.1 Lexical Bundle Use by L2S Writers across Disciplines

Table 23 shows the top 20 bundles and their frequencies (F) in each of the eight disciplines for L2S writers.

Yellow shading indicates items shared between 5-7 disciplines; blue shading shows those families shared between 2-4 disciplines, and all other items were unique to one discipline. Due to the relatively small size of the sub-corpora, only 16 bundles were identified in Social Work, while none could be identified in Physiotherapy.

Table 23-20 most frequently occurring 4-word bundles for L2S writers by discipline

Advertising	F	Education	F	Fashion	F	Forensic Psychology	F
Bundle	F	Bundle	F	Bundle	F	Bundle	F
it is important to	11	students with special needs	49	as can be seen	15	that Mr Taylor has	9
on the other hand	10	the ministry of education	34	can be seen from	14	in the case of	7
one of the most	8	will be able to	23	can be seen in	13	in an attempt to	7
are more likely to	8	of inclusive education in	23	when it comes to	12	as well as the	7
the last couple of	6	for children with SEN	23	at the same time	11	it is important to	6
in the last couple	6	on the other hand	20	for the reason that	10	in England and Wales	6
gaps have been identified	6	children with special needs	20	turns out to be	8	as a result of	6
the way in which	5	I will be able	17	the range will be	8	the Reid technique is	5
of the most important	5	to make sure that	13	the performance of the	8	the criminal justice system	5
more likely to be	5	it is important to	12	reasonable to think that	8	reduce the risk of	5
a critical evaluation of	5	to students with special	11	one of the most	8	police and criminal evidence	5
when it comes to	4	students with learning disabilities	11	it is reasonable to	8	on the other hand	5
to be able to	4	of children with SEN	11	is reasonable to think	8	it was found that	5
the way people think	4	children with SEN in	11	is one of the	8	it is necessary to	5
the research question is	4	will help me to	10	it is important to	6	it is crucial to	5
part of a community	4	when it comes to	10	the aim of this	5	it has been found	5
of the brand and	4	I will be using	10	it needs to be	5	it has also been	5
last couple of years	4	children with SEN are	10	in the fashion industry	5	in the Reid technique	5
it is also a	4	as a result of	10	despite the fact that	5	in comparison to the	5
is part of the	4	to support children with	9	can be divided into	5	and criminal evidence act	5

Languages	F	Physiotherapy	F	Social Work	F	TESOL	F
Bundle	F	Bundle	F	Bundle	F	Bundle	F
effective teaching and learning	41	-	-	on the other hand	9	on the other hand	78
it is important to	23	-	-	it is important to	9	at the end of	52
I was able to	23	-	-	in the case study	9	the end of the	41
English as an additional	23	-	-	for social workers to	7	is one of the	34
as an additional language	23	-	-	to the case study	5	at the same time	28
the rest of the	21	-	-	to social work practice	5	in the field of	27
of the target language	20	-	-	the service user to	5	as a foreign language	25
in the target language	20	-	-	have an impact on	5	in the target language	23
to be able to	19	-	-	is important to consider	4	to be able to	22
I would like to	19	-	-	that there is a	3	this study aims to	21
the use of the	18	-	-	that the service user	3	the purpose of the	21
one of the most	16	-	-	take into account the	3	native speakers of English	21
at the same time	15	-	-	of the social worker	3	it is important to	21
use of the target	13	-	-	of the limitations of	3	end of the course	21
use of target language	13	-	-	essay will discuss the	3	will be asked to	20
when it comes to	11	-	-	as it is a	3	one of the most	20
to effective teaching and	11	-	-	-		level of English proficiency	20
the end of the	11	-	-	-		at the beginning of	20
on the other hand	11	-	-	-		will be able to	19
as well as the	11	-	-	-		when it comes to	19

The absence of bundles in Physiotherapy means that no bundles are shared between all disciplines. Highest and lowest sharing by item number were in TESOL (10 shared) and Social Work (two shared). Only three bundles were common to between 5-7 disciplines, and two of these - *it is important to* and *on the other hand* - were also the two most common bundles in the corpus as a whole. Items shared between 2-4 groups included bundles such as *one of the most, to be able to, at the same time*, and *the end of the*.

There were notable differences between disciplines in terms of the potential subject specificity of unshared bundles. While not perhaps a precise indication, if those unshared bundles that are inarguably *not* subject specific (examples such as *despite the fact that, this study aims to*, and *it has been found*) are excluded, the remaining unshared item frequencies can provide a figure for the percentage of unshared item usage that *may* represent subject-specific bundles. Education and Languages showed by far the highest level of subject specificity in unshared bundles, at 11 items or around 81%, and seven items or around 63% respectively. This figure was between 25-36% for Forensic Psychology, Social Work and TESOL, and only 4-6% for Fashion and Advertising.

Bundle sharing expressed as a proportion of total top 20 usage for each discipline is presented in Table 24. The absence of bundles in Physiotherapy meant that no bundles were shared between all disciplines. Only Social work, Education and Advertising showed higher sharing in the 5-7 than in the 2-4 category. TESOL showed the highest level of sharing overall (55.15%), with Forensic Psychology lowest at only 21.21% of total top 20 usage. Unshared items accounted for over 60% of top 20 usage in every discipline apart from TESOL, suggesting that bundle usage among these writers does exhibit disciplinary variation.

Table 24-Top 20 4-word bundle sharing for L2S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 20 usage	111/100	337/100	170/100	113/100	362/100	-	79/100	553/100
Shared between all disciplines frequency / % of top 20 usage	0/0	0/0	0/0	0/0	0/0	-	0/0	0/0
Shared between 5-7 disciplines frequency / % of top 20 usage	25/22.52	42/12.46	18/10.59	11/9.74	45/12.43	-	18/22.79	118/21.34
Shared between 2-4 disciplines frequency / % of top 20 usage	12/10.81	33/9.79	27/15.88	13/11.50	92/25.41	-	0/0	187/33.81
Total shared % of top 20 usage	33.33	22.25	26.47	21.24	37.85	-	22.78	55.15
Unshared frequency / % of top 20 usage	74/66.67	262/77.75	125/73.53	89/78.76	225/62.15	-	61/77.22	248/44.85

4.3.2.2 Lexical Bundle Use by L1S Writers across Disciplines

Table 25 shows the top 20 bundles and their frequencies for L1S writers in each discipline.

Only 15 bundles met the criteria in Physiotherapy in this writer group. Item sharing is

indicated using the same colour system as in previous tables, with the addition of a white

background for items shared between all eight disciplines.

Table 25-20 most frequently occurring 4-word bundles for L1S writers by discipline

Advertising		Education		Fashion		Forensic Psychology	
Bundle	F	Bundle	F	Bundle	F	Bundle	F
the representation of women	31	for children with SEN	15	it is important to	**28**	the criminal justice system	83
representation of women in	30	with special educational needs	12	can be seen in	22	vulnerable victims and witnesses	56
of the male body	16	to be able to	11	as well as the	21	it is important to	53
it is important to	**16**	of children with SEN	10	this is due to	19	present and highly relevant	42
on the other hand	15	it is important to	**10**	the range will be	15	as a result of	42
the tone of voice	14	working with children with	8	it could be argued	15	in England and Wales	41
the gender of the	12	with children with SEN	8	is one of the	15	are more likely to	41
of women in advertising	12	when it comes to	8	at the end of	15	the way in which	36
as well as the	12	the rest of the	8	in line with the	14	relationship with his mother	36
a wide range of	12	special educational needs and	8	the end of the	13	and criminal evidence act	36
portrayal of women in	11	in order to achieve	8	the cost of the	12	is likely to be	34
on the representation of	11	I am going to	8	it is also important	12	within the criminal justice	31
when it comes to	10	children with SEN to	8	is due to the	12	it was found that	31
to the target audience	10	the other children in	7	in the fashion industry	12	Mr Taylor has a	29
one of the most	10	in a mainstream school	7	could be argued that	12	has been found to	29
John Lewis is a	10	children with SEN in	7	as well as a	12	at the time of	29
in the united states	10	with the other children	6	as can be seen	12	were more likely to	28
will be able to	9	with the aim of	6	within the fashion industry	11	have been found to	28
that there is a	9	to meet the needs	6	when it comes to	11	it is likely that	27
of gender in advertising	9	the needs of all	6	the quality of the	11	research has found that	26

Languages		Physiotherapy		Social Work		TESOL	
Bundle	F	Bundle	F	Bundle	F	Bundle	F
effective teaching and learning	58	it could be argued	7	in the case study	33	will be able to	63
the rest of the	24	quality and compassionate care	6	it is important to	**26**	at the end of	49
of the target language	24	it is important to	**6**	the service user to	23	to be able to	46
use of the target	19	high quality and compassionate	6	for social workers to	18	the end of the	46
rest of the class	19	the NHS leadership academy	4	the social worker and	17	when it comes to	36
in the target language	19	to the lack of	3	of the life course	17	the results of the	33
in the MFL classroom	19	the national health service	3	at the centre of	16	as a result of	32
I was able to	18	staff adopt leadership roles	3	to the service user	15	will be used to	31
use of target language	17	national health service NHS	3	the social worker to	15	it is important to	30
to effective teaching and	16	focused on individual leader	3	the service user and	15	in the form of	24
English as an additional	15	due to the lack	3	are more likely to	15	in the context of	24
as well as the	15	could be argued that	3	a family group conference	15	through the use of	23
as an additional language	15	as part of a	3	to the case study	14	students will be able	22
my block a placement	12	and quality of care	3	the solution focussed approach	13	is one of the	22
the use of target	11	all staff adopt leadership	3	on the other hand	13	in the field of	22
the use of the	10	-		in social work practice	13	of the English language	21
it is important to	**10**	-		to social work practice	12	the purpose of this	20
at the start of	10	-		the service user is	12	in the case of	20
teaching and learning in	9	-		the life course is	12	in line with the	20
of effective teaching and	9	-		from the case study	12	as well as the	20

Among the top 20, *It is important to* was the only bundle shared among all eight disciplines, and widespread sharing was limited compared to the L2S group; no bundles were shared between 5-7 disciplines in this group, and four of the eight disciplines had only two items shared between 2-4 disciplines. In terms of the potential subject specificity of the unshared items, levels were generally higher than amongst L2S writers. Languages and Physiotherapy showed the highest levels (14 items/~87% and nine items/~79% respectively), with Education and Social Work both demonstrating between 60-70% specificity among unshared items. The lowest potential subject specificity was found in TESOL, at two items/~18% of unshared usage.

Bundle sharing as a proportion of total top 20 usage for each discipline is given in Table 26. Unshared items make up by far the largest proportion (over 70%) of top 20 usage in six of the eight disciplines. This points towards significant disciplinary variation in L1S bundle use. The exceptions were Fashion and TESOL, in which 2-4 discipline sharing accounted for 39.46% and 55.30% respectively. Overall sharing was highest in TESOL (60.27%) and lowest in Languages (14.04%).

Table 26-Top 20 4-word bundle sharing for L1S writers across disciplines.

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 20 usage	269/100	167/100	294/100	758/100	349/100	59/100	326/100	604/100
Shared between all disciplines frequency / % of top 20 usage	16/5.95	10/5.99	28/9.52	53/6.99	10/2.86	6/10.17	26/7.98	30/4.97
Shared between 5-7 disciplines frequency / % of top 20 usage	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Shared between 2-4 disciplines frequency / % of top 20 usage	46/17.10	27/16.17	116/39.46	83/10.95	39/11.18	10/16.95	28/8,59	334/55.30
Total shared % of top 20 usage	23.05	22.16	48.98	17.94	14.04	27.12	16.56	60.27
Unshared frequency / % of top 20 usage	207/76.95	130/77.84	150/51.02	622/82.06	300/85.96	43/72.88	272/83.44	240/39.73

4.3.2.3 Lexical Bundle Use by Published writers across Disciplines

The top 20 bundles and frequencies for Published writers in each discipline are shown in Table

27. Item sharing is indicated by the same colour system utilised in previous tables.

Table 27-20 most frequently occurring 4-word bundles for Published writers by discipline

Advertising	F	Education	F	Fashion	F	Forensic Psychology	F
Bundle		Bundle		Bundle		Bundle	
in the context of	71	in the context of	37	in the context of	49	were more likely to	34
are more likely to	47	the ministry of education	32	on the other hand	33	more likely to be	30
on the other hand	40	the ways in which	28	at the same time	29	are more likely to	27
the extent to which	34	of students with SEN	27	per cent of the	28	the extent to which	26
attitude toward the brand	31	of children with disabilities	24	it is important to	28	the presence of a	24
in the case of	29	of students with disabilities	23	is positively related to	27	as a result of	22
one of the most	28	at the same time	23	the purpose of this	26	the purpose of this	21
the main effect of	24	at the end of	22	in the case of	26	it is important to	21
attitude toward the ad	22	students with special needs	21	the results of the	24	in the context of	21
the purpose of this	21	of inclusive education in	21	one of the most	24	in the control group	20
is one of the	21	the number of students	20	as well as the	24	per cent of the	18
at the same time	21	as well as the	20	on the basis of	23	in the current study	17
the relationship between the	20	it is important to	19	the degree to which	22	at the time of	17
as a result of	20	implementation of inclusive education	19	are more likely to	22	it is possible that	16
a higher level of	18	on the basis of	18	the relationship between brand	21	truth tellers and liars	15
the results of the	17	in the United States	18	the nature of the	21	there was no significant	15
as well as the	17	for children with disabilities	18	is one of the	21	in line with the	15
the mediating role of	16	to the development of	17	the extent to which	19	this paper is to	14
respondents were asked to	16	their typically developing peers	17	on the one hand	18	of the current study	14
is positively related to	16	on the other hand	17	purpose of this paper	17	be more likely to	14
languages	f	physiotherapy	f	social work	f	TESOL	F
bundle		bundle		bundle		bundle	
at the same time	46	of this study was	29	child abuse and neglect	42	the extent to which	40
in the context of	43	this study was to	26	in the context of	41	on the other hand	40
on the other hand	39	in the intervention group	25	the extent to which	31	on the basis of	39
the extent to which	32	times more likely to	18	it is important to	31	it is important to	38
it is important to	32	in patients with chronic	18	in relation to the	29	the end of the	37
the use of the	28	this is the first	17	as a result of	26	in the United States	37
in the case of	27	the purpose of this	15	on the other hand	25	in the context of	36
in relation to the	27	see online supplementary material	15	in the United States	24	in the present study	34
as well as the	27	has been shown to	15	at the same time	24	at the same time	31
the ways in which	26	purpose of this study	14	as well as the	24	as a foreign language	30
the end of the	24	the end of the	13	of social work in	20	English as a foreign	29
at the beginning of	21	body mass index BMI	13	are more likely to	20	the use of the	28
of the target language	20	is the first study	12	to be able to	18	in the form of	28
in the process of	20	at the time of	12	the development of a	18	at the end of	28
in terms of the	19	at the end of	12	to the development of	17	the ways in which	25
as a foreign language	19	the results of this	11	more likely to be	17	as a result of	24
the results of the	18	it is possible that	11	of social work practice	16	the start of the	23
in the target language	18	in the present study	11	in social work practice	16	as well as the	23
the language of the	17	as a result of	11	at the time of	16	at the beginning of	22
on the one hand	17	with the exception of	10	of child abuse and	15	the beginning of the	21

Among Published writers, there were no top 20 bundles shared between every discipline, although overall sharing was extensive, particularly in Fashion, where only three bundles were unique to the discipline. Widely shared bundles in this writer group included *as a result of, in the context of, on the other hand*, and *it is important to*. Physiotherapy had the most unshared items overall (13 out of 20). Subject specificity in unshared bundles again varied widely, ranging from 10 items/100% in Education, to one item or just over 8% in Forensic Psychology. In Physiotherapy, the discipline with the highest proportion of unshared items, only just over 25% of the unshared usage represented what could be considered subject specific bundles. This variation suggests that, in addition to direct subject specificity, there may be other factors that contribute to disciplinary variation, perhaps including established norms of the discourse communities in terms of bundle form, function and structure, as suggested by Hyland (2008a).

Item sharing as a proportion of top 20 usage is detailed in Table 28. The disciplines were split in terms of whether sharing in the 5-7, or 2-4 category constituted a larger proportion of top 20 usage. Advertising, Education, Languages, and Social Work shared more in the 5-7 category, while for the remaining disciplines, sharing with 2-4 others was more common. Total sharing varied substantially – from a minimum of 27.60% in Physiotherapy, to 87.25% in Fashion.

Table 28-Top 20 4-word bundle sharing for Published writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 20 usage	529/100	441/100	502/100	401/100	520/100	308/100	470/100	613/100
Shared between all disciplines frequency / % of top 20 usage	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
Shared between 5-7 disciplines frequency / % of top 20 usage	203/38.37	116/26.30	182/36.26	90/22.44	219/42.12	11/3.57	202/42.98	232/37.85
Shared between 2-4 disciplines frequency / % of top 20 usage	179/33.84	103/23.36	256/50.99	129/32.17	188/36.15	74/24.03	123/26.17	250/40.78
Total shared % of top 20 usage	72.21	49.66	87.25	54.61	78.27	27.60	69.15	78.63
Unshared frequency / % of top 20 usage	147/27.79	222/50.34	64/12.75	182/45.39	113/21.73	223/72.40	145/30.85	131/21.37

4.3.2.4 Summary Comparison of Writer Groups across Disciplines

Table 29 shows total sharing as a percentage of top 20 usage for the three writer groups across disciplines.

Table 29-Total sharing (% of total top 20 usage) for L2S, L1S and Published writers

		Discipline							
	Writer Group	Advertising	Education	Fashion	Forensic Psy	Languages	Physiotherapy	Social Work	TESOL
Total Sharing %	L2S	33.33	22.25	26.47	21.24	37.85	-	22.78	55.15
	L1S	23.05	22.16	48.98	17.94	14.04	27.12	16.56	60.27
	Published	72.21	49.66	87.25	54.61	78.27	27.60	69.15	78.63

Among both L2S and L1S writers, the highest proportion of item sharing was found in TESOL, although among Published writers this was in Fashion. In all disciplines, sharing was highest among Published writers. Considering RQ1, this may indicate that the most commonly used bundles in this writer group tend towards less subject specificity than among L2S and L1S writers. It may also be the case that as experienced academic writers, the Published group have a high level of familiarity with the lexical bundles commonly used in academic writing, whereas for L2S writers in particular in this study (given the generally high levels of unshared usage and relatively lower levels of potential subject specificity in the majority of the disciplines compared to their L1 counterparts) a lack of such familiarity leads to less consistent usage across disciplines.

4.3.3 Lexical Hedging

L2S writers used a total of 5683 hedges (73 of 91), with an average rate across the disciplines of 9.60 hedges per 1000 words. For the L1S group, a total of 12,356 hedges (78 of 91) were used, giving an average across disciplines of 10.45 hedges per 1000 words. Finally, the

Published writers used a total of 23,747 (85 of 91) hedges, with an across-disciplines average of 10.84 per 1000 words. Hedges per 1000 words for each writer group in each discipline are given below in Table 30, along with overall average rates for each discipline.

Table 30-Hedges per 1000 words for writer groups in each discipline

	Advertising	Education	Fashion	Forensic Psychology	Languages	Physiotherapy	Social Work	TESOL
L2S	9.26	7.97	7.29	10.67	8.57	11.08	12.35	9.63
L1S	8.63	9.91	7.37	14.24	10.19	11.12	11.50	10.63
Published	11.10	10.39	9.72	13.10	10.88	8.78	11.26	11.45
Disc. Av	9.66	9.42	8.13	12.67	9.88	10.33	11.7	10.57

L2S writers showed their highest rate of hedging in Social Work, and lowest in Fashion. For L1S writers, Forensic Psychology had the highest rate, while Fashion once again demonstrated the lowest. Published writers showed similarity with their L1S counterparts, in that they too had the highest rate in Forensic Psychology, although their lowest rate was found in Physiotherapy. Overall, taking all three writer groups into account, the highest and lowest average rates of hedging were found in Forensic Psychology and Fashion respectively.

The disciplines also varied in terms of which writer group within them hedged more or less frequently. In Advertising, Education, Fashion, Languages and TESOL, Published writers were the group with the highest rate of hedging, while in Forensic Psychology and Physiotherapy the highest rate was seen in the L1S group. Social work was the only discipline in which L2S writers hedged more frequently than their counterparts in the other groups.

The following sections will present results for the hedging analyses individual writer groups across the disciplines.

4.3.3.1 Lexical Hedge Use by L2S Writers across Disciplines

Table 31 shows the top 10 lexical hedges, along with frequency (F) and percentage of total usage for L2S writers in each discipline. Colour coding for sharing as in previous tables.

Table 31-10 Most frequently occurring lexical hedges for L2S writers by discipline

Advertising	F	%	Education	F	%	Fashion	F	%	Forensic Psychology	F	%
could	88	19.13	may	120	16.97	could	72	18.56	may	69	16.67
often	50	10.87	might	92	13.01	may	29	7.47	could	44	10.63
might	37	8.04	could	61	8.63	seem-	26	6.70	suggest-	39	9.42
likely	33	7.17	often	30	4.24	estimate-	23	5.93	likely	25	6.04
suggest-	27	5.87	tend- to	28	3.96	suggest-	22	5.67	indicate-	24	5.80
argue-	24	5.22	claim-	26	3.68	might	19	4.90	tend- to	20	4.83
may	19	4.13	mainly	25	3.54	likely	16	4.12	might	15	3.62
seem-	16	3.18	argue-	24	3.39	claim	14	3.61	often	14	3.38
possible	14	3.04	indicate-	23	3.25	indicate-	13	3.35	argue-	13	3.14
tend- to	14	3.04	usually	21	2.97	around	10	2.58	usually	12	2.90
						quite	10	2.58	claim-	12	2.90
						roughly	10	2.58			
						usually	10	2.58			
Total %		69.69	Total %		63.64	Total %		70.63	Total %		69.33

Languages	F	%	Physiotherapy	F	%	Social Work	F	%	TESOL	F	%
may	106	12.09	may	14	19.18	could	69	29.36	might	390	15.42
might	83	9.46	suggest-	9	12.33	might	23	9.79	may	279	11.03
seem-	83	9.46	might	7	9.59	suggest-	18	7.66	could	268	10.60
often	76	8.67	argue-	7	9.59	may	15	6.38	suggest-	231	9.13
could	74	8.44	claim-	6	8.22	argue-	15	6.38	seem-	130	5.14
usually	45	5.13	often	5	6.85	appear-	14	5.96	argue-	116	4.59
suggest-	33	3.76	postulate-	5	6.85	claim-	9	3.83	indicate-	90	3.56
sometimes	32	3.65	could	3	4.11	possibly	7	2.98	often	76	3.01
in my opinion	24	2.74	seem-	3	4.11	possible-	6	2.55	claim-	76	3.01
argue-	24	2.74	feel-	2	2.74	often	6	2.55	tend- to	68	2.69
			sometimes	2	2.74	indicate-	6	2.55			
Total %		66.14	Total %		86.31	Total %		79.99	Total %		68.18

The ten most commonly used hedges constituted over 60% of total usage in every discipline, peaking at 86.31% in Physiotherapy (although this sub-corpus was of very small size). In every discipline a modal verb was the most commonly used hedge; *may* in four disciplines, *could* in

three, and *might* in one. Indeed these three modals accounted for a substantial proportion of all usage within the top ten hedges, ranging from a minimum of 38.10% in Physiotherapy, to a maximum of 60.67% in Education. Modals and lexical verbs were the predominant choice of hedging device in all eight disciplines.

 Where commonality between disciplines is concerned, it was the three modals *may*, *could* and *might* that were shared most widely, with lexical verbs such as suggest, argue, seem and claim also widespread. Often was the only non-verb hedge shared between more than four of the disciplines. The exact proportions of sharing in each category are given below in Table 32.

 Sharing between all disciplines made up the highest proportion of usage within the top ten for all disciplines with the exception of Physiotherapy, although the very low numbers in this sub-corpus make inferences difficult. Overall sharing was generally high across the disciplines, with a minimum of 80.66% of top ten usage shared in Fashion. Two disciplines, Advertising and TESOL had no unshared items among the top ten hedges.

Table 32-Top 10 lexical hedge sharing for L2S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 10 usage	322/100	450/100	274/100	287/100	580/100	63/100	188/100	1724/100
Shared between all disciplines frequency / % of top 10 usage	144/44.72	273/60.67	120/43.80	128/44.60	263/45.35	24/38.10	107/56.92	937/54.35
Shared between 5-7 disciplines frequency / % of top 10 usage	117/36.34	103/22.89	75/27.37	102/35.54	216/37.24	30/47.62	54/28.72	719/41.71
Shared between 2-4 disciplines frequency / % of top 10 usage	61/18.94	49/10.89	26/9.49	57/19.86	45/7.76	0/0	6/3.19	68/3.94
Total shared % of top 10 usage	100	94.44	80.66	100	90.35	85.71	88.83	100
Unshared frequency / % of top 10 usage	0/0	25/5.56	5/19.34	0/0	56/9.65	9/14.29	21/11.17	0/0

4.3.3.2 Lexical Hedge Use by L1S Writers across Disciplines

Top 10 lexical hedges, with frequency (F) and percentage of total usage for L2S writers in each discipline are given in Table 33. Colour coding as in previous tables.

Advertising	F	%	Education	F	%	Fashion	F	%	Forensic Psychology	F	%
Hedge			Hedge			Hedge			Hedge		
could	178	14.66	may	112	22.67	could	186	16.67	may	935	23.76
may	140	11.53	often	62	12.55	may	164	14.77	suggest-	577	14.66
suggest-	131	10.79	could	52	10.53	suggest-	88	7.93	could	460	11.69
often	90	7.41	appear-	30	6.07	indicate-	62	5.59	likely	407	10.34
indicate-	60	4.94	argue-	30	6.07	likely	61	5.50	often	221	5.62
argue- -	58	4.78	feel-	29	5.87	estimate-	56	5.05	indicate-	159	4.04
likely -	48	3.95	suggest-	27	5.47	often	43	3.87	seem-	135	3.34
feel-	41	3.78	likely	19	3.85	seem-	34	3.06	appear-	122	3.10
seem-	39	3.21	seem-	13	2.63	around	31	2.79	feel-	121	3.07
generally	31	2.55	usually	9	1.82	typically	28	2.52	argue-	103	2.62
Total %		67.60	Total %		77.53	Total %		67.75	Total %		82.24

Languages	F	%	Physiotherapy	F	%	Social Work	F	%	TESOL	F	%
Hedge			Hedge			Hedge			Hedge		
may	122	14.56	suggest-	51	20.40	may	249	27.01	may	585	16.28
often	82	9.79	may	46	18.40	could	164	17.79	suggest-	524	14.58
could	81	9.67	could	44	17.60	suggest-	149	16.16	could	399	11.10
feel-	62	7.40	often	18	7.20	argue-	67	7.27	often	214	5.96
suggest-	52	6.21	argue-	14	5.60	likely	50	5.42	argue-	123	3.42
seem-	46	5.49	appear-	10	4.00	often	33	3.58	seem-	119	3.41
perhaps	32	3.82	tend- to	6	2.40	feel-	23	2.49	likely	101	2.81
likely	29	3.46	assume-	6	2.40	sometimes	19	2.06	appear-	101	2.81
argue-	29	3.46	generally	5	2.00	appear-	19	2.06	indicate-	89	2.48
appear-	26	3.10	likely	5	2.00	might	18	1.95	possible	87	2.42
			indicate-	5	2.00				feel-	87	2.42
Total %		66.96	Total %		84.00	Total %		85.79	Total %		67.69

The proportion of total usage represented by the top ten hedges ranged from 66.96% in Languages, to 85.79% in Social Work. Among L1S writers the modals *may* and *could* were the most common hedges (*may* in five disciplines and *could* in two). Only Physiotherapy differed, with the lexical verb *suggest.* Interestingly, the modal *might* was absent from the top ten in all but one discipline – Social Work. Perhaps because of this, modals accounted for generally less of the top ten usage in this group than they did for L2S writers, the only exceptions being Physiotherapy and Fashion, where levels were slightly higher among the L1S group.

Top 10 sharing was more extensive among L1S writers, with not only the two modals, but also the lexical verb *suggest*, the adverb *often*, and the adverb/adjective *likely* being common to all eight disciplines. The shared items common to 5-7 disciplines were all lexical verbs, while the only item in the shared between 2-4 category was the adverb *generally*. Sharing proportions are given in Table 34.

Table 34-Top 10 lexical hedge sharing for L1S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 10 usage	816/100	383/100	753/100	3240/100	561/100	210/100	791/100	2429/100
Shared between all disciplines frequency / % of top 10 usage	587/71.94	272/71.02	542/71.98	2600/80.25	366/65.24	164/78.10	645/81.54	1823/75.05
Shared between 5-7 disciplines frequency / % of top 10 usage	198/24.27	102/26.63	96/12.75	640/19.75	163/29.06	29/13.81	109/13.78	519/21.37
Shared between 2-4 disciplines frequency / % of top 10 usage	31/3.80	0/0	0/0	0/0	0/0	5/2.38	0/0	0/0
Total shared % of top 10 usage	100	97.65	84.73	100	94.30	94.29	95.32	96.42
Unshared frequency / % of top 10 usage	0/0	9/2.35	115/15.23	0/0	32/5.70	12/5.71	37/4.68	87/3.58

Sharing across all eight disciplines was by far the highest proportion of usage within the top ten in all disciplines (min. 65.24% in Languages). Total sharing was higher when compared to

L2S writers in seven of the eight disciplines, the exception being TESOL, in which one item was

unshared, thus reducing total sharing from 100% to 96.42%.

4.3.3.3 Lexical Hedge Use by Published writers across Disciplines

Table 35 indicates the top 10 lexical hedges for Published writers in each discipline. Hedges are

given with frequencies (F) and percentage of total usage. Colour coding for sharing as in

previous tables.

Table 35-10 most frequently occurring lexical hedges for Published writers by discipline

Advertising			Education			Fashion			Forensic Psychology		
Hedge	F	%	Hedge	F	%	Hedge	F	%	Hedge	F	%
may	571	17.03	may	345	12.49	may	559	18.92	may	604	21.11
suggest-	333	9.93	suggest-	238	8.61	suggest-	286	9.68	suggest-	275	9.61
could	252	7.52	could	200	7.24	could	185	6.26	likely	259	9.05
indicate-	248	7.40	often	189	6.84	often	176	5.96	indicate-	189	6.61
likely	246	7.34	indicate-	189	6.84	likely	175	5.92	could	183	6.40
might	175	5.22	might	165	5.97	indicate-	169	5.72	might	139	4.86
often	150	4.47	argue-	131	4.74	might	124	4.20	often	130	4.54
tend to-	106	3.16	likely	86	3.11	estimate-	123	4.16	appear-	76	2.66
seem-	104	3.10	seem-	101	3.66	argue-	94	3.18	possible	67	2.34
in general	97	2.89	possible	75	2.71	tend- to	79	2.67	generally	57	1.99
Total %		68.06	*Total %*		62.21	Total %		66.67	Total %		69.17
Languages			Physiotherapy			Social Work			TESOL		
Hedge	F	%	Hedge	F	%	Hedge	F	%	Hedge	F	%
may	529	16.92	may	305	20.76	may	528	17.50	may	638	15.18
suggest-	294	9.40	indicate-	149	10.14	suggest-	257	8.52	suggest-	367	8.73
could	171	5.47	could	148	10.07	might	232	7.69	indicate-	265	6.31
might	168	5.37	suggest-	141	9.60	likely	192	6.36	often	245	5.83
indicate-	164	5.24	likely	98	6.67	indicate-	187	6.20	might	245	5.83
often	157	5.02	estimate-	80	5.45	often	182	6.03	likely	192	4.57
seem-	157	5.02	might	72	4.90	could	140	4.64	seem-	168	4.00
argue-	141	4.19	often	49	3.34	argue-	121	4.01	could	167	3.97
appear-	119	3.81	approximately	45	3.06	tend- to	84	2.78	argue-	148	3.52
likely	94	3.01	possible	39	2.65	seem-	75	2.49	appear-	109	2.59
Total %		63.45	*Total %*		76.64	Total %		66.22	Total %		60.53

The ten most commonly used hedges accounted for between 60.53% (TESOL) and 76.64%

(Physiotherapy) of total hedge use. *May* was the most common hedge in every discipline.

Indeed, seven of the disciplines shared the same top two hedges - *may and suggest*, while four shared the top three - *may, suggest* and *could.* Modals accounted for between 41.27% (TESOL) and 46.79% (Forensic Psychology) of top ten usage, and the modal *might* reappeared in every discipline. Along with the three modals, the lexical verbs *suggest* and *indicate*, the adverb *often*, and the adverb/adjective *likely* were shared between all disciplines. Verbs were the most common hedging device once again. A detailed picture of item sharing by frequency is given in Table 36.

Table 36-Top 10 lexical hedge sharing for Published writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 10 usage	2282/100	1719/100	1970/100	1979/100	1994/100	1126/100	1998/100	2544/100
Shared between all disciplines frequency / % of top 10 usage	1975/86.55	1412/82.14	1674/84.98	1779/89.89	1557/79.09	962/85.44	1718/85.99	2119/83.29
Shared between 5-7 disciplines frequency / % of top 10 usage	104/4.56	232/13.50	94/4.77	0/0	298/14.95	0/0	196/9.81	216/17.42
Shared between 2-4 disciplines frequency / % of top 10 usage	106/4.65	75/4.36	202/10.25	143/7.23	119/5.97	39/3.46	84/4.20	109/4.29
Total shared % of top 10 usage	95.75	100	100	97.12	100	88.90	100	100
Unshared frequency / % of top 10 usage	97/4.25	0/0	0/0	57/2.88	0/0	125/11.10	0/0	0/0

Sharing between all eight disciplines was the highest proportion of top ten usage, and was higher among Published writers than among the L2S or L1S groups in every discipline. Total sharing was also very high in this group, with five disciplines having no unshared items. These results suggest that among Published writers there is considerable uniformity in the most commonly used hedges across disciplines.

4.3.3.4 Summary Comparison of Writer Groups across Disciplines.

Table 37, below, summarises the total sharing for each writer group across disciplines (as a percentage of total top 10 usage) as well as the percentage of top ten usage accounted for by modal verbs.

Table 37-Total sharing and modals (% of top 10 usage) for L2S, L1S and Published writers

	Writer Group	Discipline							
		Advertising	Education	Fashion	Forensic Psy	Languages	Physiotherapy	Social Work	TESOL
Total Sharing %	L2S	100	94.44	80.66	100	90.35	85.71	88.83	100
	L1S	100	97.65	84.73	100	94.30	94.29	95.32	96.42
	Published	95.75	100	100	97.12	100	88.90	100	100
Modals %	L2S	44.72	60.67	43.80	44.60	45.35	38.10	56.92	54.35
	L1S	38.97	42.82	44.60	43.06	36.19	42.86	54.49	40.51
	Published	43.73	41.30	44.06	46.79	43.53	46.63	45.05	41.27

Among both L2S and L1S writers, total sharing was lowest in Fashion. Sharing was highest in Advertising, Forensic Psychology and TESOL (L2S), and Advertising, Forensic Psychology and Education (L1S). For Published writers, Physiotherapy saw the lowest level of sharing, while five disciplines were at 100% of top ten usage. In terms of which writer group showed the highest or lowest sharing, in five disciplines - Education, Fashion, Languages, Physiotherapy, and Social Work, sharing was lowest among L2S writers. Published writers also demonstrated

the highest levels of sharing in four of these disciplines (Education, Fashion, Languages, and Social Work). In terms of RQ1, disciplinary variation, while generally limited, would seem to be more extensive among L2S writers. There was no particular pattern to modal use either in terms of disciplines being higher or lower across different writer groups, or of writer groups being higher or lower across different disciplines. In Education and TESOL, modal use was notably higher among L2S writers compared to the other writer groups, but levels of modal usage were generally fairly consistent.

4.3.4 Citations

The following sections will present results for the citation analyses of the various sub-corpora. Firstly, each writer group will be compared across the disciplines, and this will be followed by the results for the three writer groups compared within each discipline.

4.3.4.1 Citation Use by L2S Writers

Overall citation use by L2S writers in each discipline is shown below in Table 38. This details citation totals, rates of citation use per 1000 words, non-integral and integral citation use as a percentage of total citations, and the various forms of integral citation, as a percentage of total integral citation use.

Table 38-Overall citation use by L2S writers across disciplines

	Advertising	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total citations	308	948	433	489	429	57	354	3262
Non-integral citations	208	744	374	322	250	30	276	1961
Integral citations	100	204	59	167	179	27	78	1301
Citations/1000 words	6.20	10.69	8.13	12.60	4.19	8.65	18.60	12.43
Non-integral citations %	67.53	78.48	86.37	65.85	58.28	52.63	77.97	60.12
Integral citations %	32.47	21.52	13.63	34.15	41.72	47.37	22.03	39.88
Non-Reporting integral citations %	18.00	12.74	30.50	8.98	11.17	3.70	8.97	20.13
According to %	29.00	20.59	35.59	7.78	11.73	7.40	3.85	13.07
Discourse act verbs %	33.00	46.57	20.34	32.93	54.75	51.85	65.38	46.66
Research act verbs %	17.00	12.75	10.17	41.92	15.64	37.04	21.79	15.22
Cognition act verbs %	3.00	7.35	3.39	8.38	6.70	0.00	0.00	4.77
Ambiguous %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15

The rates of citation use per 1000 words differed considerably across the disciplines, with the highest and lowest rates seen in Social Work (18.60) and Languages (4.19) respectively. Non-integral citations were more common than integral citations in every discipline, although the difference in percentage usage was relatively small in Physiotherapy (52.63% and 47.37%) and at its most extensive in Fashion (86.37% and 13.63%). In most disciplines non-integral citations constituted at least 60% of total citation use. Literature suggests that non-integral citations place more emphasis on the content being cited than on the author (Hyland, 1999b), although whether this was the main motivation for their use here is unclear. Considering integral citations themselves, there was wide disciplinary variation in terms of the frequency of the various forms. The preposition *according to* constituted a higher proportion of usage in Advertising, Education, and Fashion than in the other disciplines, while its use was particularly low in Social Work, Physiotherapy and Forensic Psychology; again, why this should be so is uncertain. Of the three types, discourse act verbs dominated in all but Forensic Psychology (where research act verbs were more common), and cognition act verbs represented the smallest proportion of integral citation use in every discipline.

Turning to the specific verbs used themselves, and how they were shared across the disciplines, Table 39 shows the ten most frequently used reporting verbs for integral citations by L2S writers in each discipline, with frequencies (F) and percentage of total integral citation use (colour coding as previous tables). Coverage provided by the ten most frequently used reporting structures varied across the disciplines, from a low of 42.42% in TESOL, to 96.29% in Physiotherapy, although this was probably influenced by the very different sizes of some of the sub-corpora in this comparison, and by the large number of identical frequencies in Physiotherapy. Only three top 10 items, *according to, suggest* and *argue*, were common to all eight disciplines (these were also among the five most frequently used items in the corpus as a whole), while only two were shared between 5-7 disciplines. Among widely shared items, there was a clear tendency towards discourse act verbs. Education showed the most commonality with other disciplines, with no unshared items among the top 10. Conversely, in terms of item number, Fashion was the discipline with most unshared items.

Table 39-10 most frequently occurring reporting verbs for L2S writers by discipline

Advertising			Education			Fashion			Forensic Psychology		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	29	29.00	according to	42	20.59	according to	21	35.59	find	40	23.95
state	7	7.00	state	17	8.33	state	3	5.08	according to	13	7.78
suggest	4	4.00	claim	12	5.88	suggest	2	3.39	suggest	9	5.39
argue	3	3.00	suggest	7	3.43	report	2	3.39	support	8	4.79
show	3	3.00	describe	6	2.94	argue	2	3.39	discover	7	4.19
define	3	3.00	argue	6	2.94	say	1	1.69	*conduct*	7	4.19
explain	3	3.00	support	5	2.45	predict	1	1.69	argue	5	2.99
find out	3	3.00	mention	5	2.45	mention	1	1.69	emphasise	5	2.99
point out	2	2.00	define	5	2.45	list	1	1.69	report -	5	2.99
imply	2	2.00	refer to	4	1.96	introduce	1	1.69	show	5	2.99
believe	2	2.00	explain	4	1.96	conduct	1	1.69			
						campaign	1	1.69			
						advise	1	1.69			
						sustain	1	1.69			
						succeed	1	1.69			
						take into account	1	1.69			
Total %		61.00	Total %		55.38	Total %		69.43	Total %		62.25

Languages			Physiotherapy			Social Work			TESOL		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	21	11.73	claim	3	11.11	suggest	8	10.26	according to	170	13.07
argue	14	7.82	according to	2	7.41	define	7	8.97	suggest	82	6.30
state	12	6.70	describe	2	7.41	highlight	6	7.69	state	72	5.53
point out	10	5.59	explore	2	7.41	argue	5	6.41	argue	54	4.15
mention	8	4.47	find	2	7.41	*describe*	4	5.13	mention	44	3.38
suggest	6	3.35	study	2	7.41	develop	4	5.13	claim	36	2.77
explain	6	3.35	state	2	7.41	state	4	5.13	*find*	31	2.38
show -	5	2.79	report	2	7.41	according to	3	3.85	point out	22	1.69
affirm	4	2.23	identify	2	7.41	*find*	3	3.85	cite	21	1.61
establish	3	1.68	suggest	1	3.70	explain	3	3.85	refer to	20	1.54
carry out	3	1.68	show	1	3.70	point out	3	3.85			
write	3	1.68	point out	1	3.70						
			focus on	1	3.70						
			argue	1	3.70						
			alert	1	3.70						
			advocate	1	3.70						
Total %		53.07	Total %		96.29	Total %		64.12	Total %		42.42

Table 40 details L2S item sharing across disciplines in terms of percentages of total usage within the ten most frequent reporting verbs.

Table 40-Top 10 reporting verb sharing for L2S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 10 usage	61/100	113/100	41/100	104/100	95/100	26/100	50/100	552/100
Shared between all disciplines frequency / % of top 10 usage	36/59.02	55/48.67	25/60.98	27/25.96	41/43.16	4/15.39	16/32.00	306/55.44
Shared between 5-7 disciplines frequency / % of top 10 usage	9/14.75	17/15.04	3/7.32	0/0	22/23.16	2/7.69	7/14.00	94/17.03
Shared between 2-4 disciplines frequency / % of top 10 usage	9/14.75	41/36.28	4/9.76	65/62.50	19/20.00	10/38.46	17/34.00	131/23.73
Total shared % of top 10 usage	88.52	100	78.05	88.46	86.32	61.54	80.00	96.17
Unshared frequency / % of top 10 usage	7/11.48	0/0	9/21.95	12/11.54	13/13.68	10/38.46	10/20.00	21/3.80

Sharing overall was generally high among L2S writers, with 61.54% of top ten usage the minimum (Physiotherapy). In most cases, the largest proportion of top 10 usage was those items shared between all eight disciplines (as these were often among the most frequently used items). The notable exceptions to this were Forensic Psychology and Physiotherapy, in which 2-4 group sharing was proportionally more common. For the former, this was also the only discipline to employ proportionally more research verbs, and the presence of verbs such as *find, conduct, discover,* and *show* may reflect this. In the case of the latter, this discipline

also showed the greatest proportion of unshared items, although given the very small size of this sub-corpus, it would be unwise to generalise from this.

4.3.4.2 Citation Use by L1S Writers

Overall citation use by L1S writers in each discipline is shown below in Table 41.

Table 41-Overall citation use by L1S writers across disciplines

	Advertising	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total citations	942	530	1295	3225	333	214	1359	2834
Non-integral citations	602	396	1148	2171	178	73	897	1217
Integral citations	340	134	147	1054	155	141	462	1617
Citations/1000 words	6.70	10.63	8.59	11.67	4.05	9.51	16.95	8.38
Non-integral citations %	63.91	74.72	88.65	67.32	53.45	34.11	66.00	42.94
Integral citations %	36.09	25.28	11.35	32.68	46.55	65.89	34.00	57.06
Non-Reporting integral citations %	14.41	8.96	8.84	13.66	6.45	4.25	11.90	18.12
According to %	4.41	0.75	20.40	1.61	1.29	0.00	4.98	4.64
Discourse act verbs %	37.94	57.46	51.02	35.86	64.52	55.32	63.85	48.86
Research act verbs %	36.18	13.43	15.65	43.74	17.42	25.53	14.07	21.89
Cognition act verbs %	7.06	19.40	3.40	5.12	9.03	14.89	4.76	6.18
Ambiguous %	0.00	0.00	0.68	0.00	1.29	0.00	0.43	0.31

As with the L2S group, the highest and lowest rates of citation per 1000 words occurred in Social Work and Languages respectively. Indeed, rates across the disciplines for the two writer groups were broadly similar. Non-integral citations once again dominated, although notable exceptions to this for L1S writers were Physiotherapy and TESOL, where integral citations were actually more common. It is unclear whether writers in these disciplines consciously made more frequent use of citations placing emphasis on authors rather than content, or whether there are other reasons for integral citations to be more frequent here. For integral citations themselves, the proportion of non-reporting forms varied across the disciplines, but was lowest in Physiotherapy and highest in TESOL. The use of *according to* was notably higher in Fashion than in any other discipline. Discourse act verbs were dominant in every discipline

other than Forensic Psychology, in which research verbs made up a larger proportion. Education was the only discipline where cognition verbs were not the lowest of the three verb types by proportion.

 The most frequently used reporting verbs and how they were shared across disciplines by L1S writers is shown in Table 42 (colour code as previous tables). Top 10 coverage as a percentage of total integral citation use ranged from around 40% in TESOL, to over 76% in Fashion. Only two top 10 reporting verbs were common to all eight disciplines – *suggest* and *find,* with *state, argue, describe* and *discuss* found in 5-7 of the disciplines; this shows the general dominance of discourse act verbs. In the majority of the disciplines, at least in terms of item number, sharing in the 2-4 category was the most common. Both Forensic Psychology and TESOL shared all of their top 10 reporting verbs.

 Table 43 details L1S item sharing across disciplines in terms of percentages of total usage within the ten most frequent reporting verbs. Total sharing was generally very high among L1S writers, at over 92% in all but Physiotherapy. Compared to L2S writers, there was a relatively more even spread between the three sharing categories, with the single largest percentage in any one category being 54.65% for 5-7 sharing in Languages. Fashion was the only discipline in which the 2-4 category represented the highest proportion, and this may be due to the relatively high proportion of top 10 usage attributable to *according to.*

Table 42-10 most frequently occurring reporting verbs for L1S writers by discipline

Advertising			Education			Fashion			Forensic Psychology		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
find	45	13.24	argue	23	17.16	state	31	21.09	find	230	21.82
state	31	9.12	state	16	11.94	according to	30	20.41	state	98	9.30
suggest	17	5.00	believe	14	10.45	report	12	8.16	suggest	67	6.36
according to	15	4.41	suggest	9	6.72	suggest	11	7.48	conduct	43	4.08
argue	11	3.24	agree	4	2.99	indicate	4	2.72	argue	26	2.47
conduct -	10	2.94	define	4	2.99	discuss	4	2.72	note	25	2.37
explain	10	2.94	find	4	2.99	define	3	2.04	describe	24	2.28
support	8	2.35	claim	3	2.24	describe	3	2.04	conclude	19	1.80
describe	7	2.06	discuss	3	2.24	find	3	2.04	identify	19	1.80
conclude	5	1.47	report	3	2.24	identify	3	2.04	support	19	1.80
say	5	1.47	say	3	2.24	note	2	1.36			
claim	5	1.47	uncover	3	2.24	outline	2	1.36			
			view	3	2.24	prove	2	1.36			
						provide	2	1.36			
Total %		49.71	Total %		68.68	Total %		76.18	Total %		54.08

Languages			Physiotherapy			Social Work			TESOL		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
state	20	12.90	suggest	30	21.28	suggest	75	16.23	suggest	178	11.01
argue	14	9.03	find	11	7.80	state	34	7.36	state	76	4.70
suggest	11	7.10	agree	9	6.38	argue	33	7.14	according to	75	4.64
describe	9	5.81	discuss	8	5.67	describe	25	5.41	find	64	3.96
highlight	6	3.87	describe	7	4.96	according to	23	4.98	note	55	3.40
support	5	3.23	support	7	4.96	find	15	3.25	argue	54	3.34
define	4	2.58	conduct	6	4.26	propose	12	2.60	propose	46	2.84
discuss	4	2.58	present	6	4.26	affirm	10	2.16	discuss	37	2.29
say	4	2.58	argue	4	2.84	highlight	9	1.95	define	35	2.16
find	3	1.94	confirm	4	2.84	note	8	1.73	describe	35	2.16
believe	3	1.94				outline	8	1.73			
advocate	3	1.94									
Total %		55.50	Total %		65.25	Total %		54.54	Total %		40.50

Table 43-Top 10 reporting verbs sharing for L1S writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 10 usage	169/100	92/100	112/100	570/100	86/100	92/100	252/100	655/100
Shared between all disciplines frequency / % of top 10 usage	62/36.69	13/14.13	14/12.50	297/52.10	14/16.28	41/44.57	90/35.71	242/36.95
Shared between 5-7 disciplines frequency / % of top 10 usage	49/28.99	39/42.39	38/33.93	148/25.97	47/54.65	19/20.65	92/36.51	202/30.84
Shared between 2-4 disciplines frequency / % of top 10 usage	48/28.40	34/36.96	52/46.43	125/21.93	22/25.58	22/23.91	60/23.81	211/32.21
Total shared % of top 10 usage	94.08	93.48	92.86	100	96.51	89.13	96.03	100
Unshared frequency / % of top 10 usage	10/5.92	6/6.52	8/7.14	0/50	3/3.49	10/10.87	10/3.97	0/0

4.3.4.3 Citation Use by Published writers

Overall citation use by Published writers in each discipline is shown below in Table 44.

Table 44-Overall citation use by Published writers across disciplines

	Advertising	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total citations	1421	1316	2499	1625	1237	1804	2171	2546
Non-integral citations	728	728	1914	1250	524	1714	1648	1612
Integral citations	693	588	585	375	713	90	523	934
Citations/1000 words	4.70	4.95	8.22	7.44	4.31	10.78	8.10	6.94
Non-integral citations %	51.23	55.32	76.59	76.92	42.36	95.01	75.91	63.32
Integral citations %	48.77	44.68	23.41	23.08	57.64	4.99	24.09	36.68
Non-Reporting integral citations %	22.51	25.00	27.01	16.27	24.82	25.56	19.50	26.02
According to %	6.20	3.74	6.32	0.53	6.03	2.22	3.63	2.14
Discourse act verbs %	33.77	41.33	35.90	37.60	39.55	27.78	43.60	36.51
Research act verbs %	32.76	24.32	27.86	40.53	25.25	42.22	28.30	31.16
Cognition act verbs %	3.03	5.27	2.39	4.80	3.37	2.22	3.82	3.53
Ambiguous %	1.73	0.34	0.51	0.27	0.98	0.00	1.15	0.64

Citation rates were highest in Physiotherapy and lowest in Languages. Non-integral citations dominated in all disciplines[3] other than Languages, which suggests a general tendency towards emphasising content rather than author in the majority of the disciplines (assuming this is something Published writers are conscious of). The proportion of non-reporting structures was broadly similar across disciplines, although slightly lower in Social Work and Forensic Psychology. *According to* was used infrequently in every discipline compared to either non-reporting forms or discourse and research act verbs. Research act verbs were dominant only in Forensic Psychology and Physiotherapy, while cognition acts were the least used by proportion in every discipline. Table 45 shows the ten most frequently used reporting verbs in each discipline, along with how they were shared (colour code as previous tables). The top ten

---

[3] A numerical superscript system was employed in the research articles in Physiotherapy, meaning that the vast majority of citations were classified as non-integral. Despite this system, a relatively small number of integral citations were nonetheless found in the text.

items accounted for between ~33% to ~53% of total integral citation usage. Three top 10 items – *find, suggest*, and *show* - were shared across all eight disciplines. Discourse act verbs were common among the most widely shared items (*suggest, argue, describe, report*, for example), while research act verbs occurred commonly in the 2-4 sharing category (*identify, conduct, develop, demonstrate, examine*). By item number, Advertising had the most unshared items.

Table 45-10 most frequently occurring reporting verbs for Published writers by discipline

Advertising			Education			Fashion			Forensic Psy		
Rep Str	F	%	Rep Str	F	%	Rep Str	F	%	Rep Str	F	%
suggest	51	7.36	argue	40	6.8	according to	37	6.32	find	40	10.67
find	48	6.93	describe	28	4.76	suggest	36	6.15	suggest	34	9.07
according to	43	6.20	suggest	27	4.59	find -	32	5.47	report	28	7.47
define	26	3.75	according to	22	3.74	argue	24	4.1	identify	19	5.07
propose	25	3.61	find	22	3.74	show	19	3.25	describe	16	4.37
develop	24	3.46	show	13	2.21	report	18	3.08	conduct	13	3.47
identify	23	3.32	report	13	2.21	develop	15	2.56	show	10	2.67
state	20	2.89	note -	12	2.04	point out	12	2.05	examine	6	1.60
note	18	2.60	identify	12	2.04	propose	12	2.05	develop	6	1.60
show	15	2.16	propose	11	1.87	use	12	2.05	conclude	6	1.60
demonstrate	14	2.02	conduct	9	1.53	describe	12	2.05			
Total %		44.30	Total %		35.53	Total %		39.13	Total %		47.59
Languages			Physiotherapy			Social Work			TESOL		
Rep Str	F	%	Rep Str	F	%	Rep Str	F	%	Rep Str	F	%
find	51	7.15	find	12	13.33	find	36	6.88	find	78	8.35
according to	43	6.04	report	6	6.67	suggest	29	5.54	argue	49	5.25
argue	37	5.19	suggest	5	5.56	argue	24	4.59	describe	36	3.85
note	25	3.51	conduct	4	4.44	note	23	4.40	show	29	3.10
suggest	20	2.81	analyse	3	3.33	according to	19	3.63	suggest	25	2.68
report	20	2.81	examine	3	3.33	identify	19	3.63	note	22	2.36
describe	19	2.66	note	3	3.33	report	17	3.25	according to	20	2.14
show	18	2.52	according to	2	2.22	show	11	2.10	report	20	2.14
observe	13	1.82	demonstrate	2	2.22	describe	10	1.91	investigate	18	1.93
propose	13	1.82	describe	2	2.22	examine	9	1.72	point out	18	1.93
			observe	2	2.22	point out	9	1.72			
			recommend	2	2.22						
			show	2	2.22						
Total %		36.33	Total %		53.31	Total %		39.37	Total %		33.73

Sharing across disciplines as a proportion of total top 10 usage is shown in Table 46. Sharing in general was very high among Published writers, with the majority of disciplines sharing over 94% of top ten usage. Advertising showed the lowest level of sharing, but even this amounted to over 85%.

Table 46-Top 10 reporting verb sharing for Published writers across disciplines

	Discipline							
	Adverting	Education	Fashion	Forensic Psy	Languages	Physio	Social Work	TESOL
Total frequency / % of top 10 usage	307/100	209/100	229/100	178/100	259/100	48/100	206/100	315/100
Shared between all disciplines frequency / % of top 10 usage	114/37.13	62/29.67	87/37.99	84/47.19	89/34.36	19/39.58	76/36.89	132/41.91
Shared between 5-7 disciplines frequency / % of top 10 usage	61/19.87	115/55.02	91/39.74	44/24.72	144/55.99	11/29.92	93/45.15	147/46.67
Shared between 2-4 disciplines frequency / % of top 10 usage	86/28.01	32/15.31	39/17.03	44/24.72	26/10.04	13/22.08	37/17.96	18/5.71
Total shared % of top 10 usage	85.02	100	94.76	96.63	100	89.58	100	94.29
Unshared frequency / % of top 10 usage	46/14.98	0/0	12/5.24	6/3.37	0/0	5/10.42	0/0	18/5.71

4.3.4.4 Summary Comparison of Writer Groups across Disciplines

For all writer groups, the rate of citations per 1000 words was lowest in languages. Non-reporting forms were highest across the disciplines among Published writers, with the exception of Fashion, in which L2S writers showed the highest proportion. In all three writer

171

groups, discourse verbs were the most frequently used in citations across the disciplines, with the notable exceptions of Forensic Psychology (in which, in all writer groups, research verbs made up the largest proportion from the three verb types), and Physiotherapy (in which, among Published writers, research act verbs were the most commonly used type). Cognition act verbs were generally the least frequently used across writers groups and disciplines, although L1S writers in Education and Physiotherapy showed an unusually high proportion of usage for these verbs. The proportions of top 10 usage that were shared by each writer group in each discipline are shown in Table 47, below.

Table 47-Total reporting verb sharing (% of total top 10 usage) for each writer group in each discipline

	Writer Group	Discipline							
		Advertising	Education	Fashion	Forensic Psy	Languages	Physiotherapy	Social Work	TESOL
Total Sharing %	L2S	88.52	100	78.05	88.46	86.32	61.54	80.00	96.17
	L1S	94.08	93.48	92.86	100	96.51	89.13	96.03	100
	Published	85.02	100	94.76	96.63	100	89.58	100	94.29

There was some consistency between disciplines in terms of which writer groups showed the highest or lowest levels of item sharing. In Social Work, Physiotherapy, Languages, and Fashion, Published writers showed the highest level of sharing, while in Education levels were equal between L2S and Published writers. L1S writers showed the highest levels in Advertising, Forensic Psychology, and TESOL. In terms of which disciplines were highest or lowest in the different writer groups, Physiotherapy showed the lowest rates among both student groups, while sharing levels in Education were highest for L2S writers, and among the highest in the Published group. Overall, in reference to RQ1, the generally high levels of item sharing across

disciplines in all three writer groups indicate that disciplinary variation in the most commonly used reporting verbs is, at least in the disciplines examined here, limited.

4.4 RQ2 - Variation across writer groups

4.4.1 In the corpus as a whole

Sections 4.4.1.1 – 4.4.1.4 present the results of comparisons between the three writer groups in terms of the four writing features across the corpus as a whole, in other words, with the eight disciplines combined to provide an overall picture of how the features were used by each writer group.

4.4.1.1 Vocabulary

4.4.1.1(i) AWL and GSL coverage

As explained in Section 4.3.1, the examination of AWL and GSL type coverage was only made possible by normalising any sub-corpora to be compared so as to account for the effects of text length. Comparisons of a single writer group across the eight disciplines, such as those performed in Sections 4.3.1.1–4.3.1.3, are possible because discipline text length is the only variable that needs to be balanced. Likewise, comparisons of the three writer groups within a single discipline, as conducted in Section 4.4.2.1, are possible, since only writer group text length need be balanced. However, the normalisation process presents a problem when it comes to comparing writer groups with disciplines combined, since this would require both writer group text length *and* discipline text length to be balanced. By way of illustration, the total word counts for each writer group in the corpus are: L2S - 620,937, L1S - 1,140, 451, and Published - 2,179, 883. While these could, in theory, be normalised for the smallest (L2S), the word counts for each discipline within these three sub-corpora vary considerably, and would thus not be balanced. This would mean that any calculation of type coverage for a given writer

group would be skewed by the different text lengths of the disciplines within that group, and as a consequence, the overall comparison of the writer groups themselves would be invalidated. For this reason it was not possible to perform combined-disciplines analysis to compare AWL and GSL type coverage across writer groups.

4.4.1.1.(ii) Frequency of AWL Vocabulary

Table 48a shows the 20 most frequent AWL families, with frequencies and cumulative percentages for the three writer groups in the corpus as a whole. It also indicates the extent of item sharing. Family sharing between writer groups is denoted in the following way: White shading indicates families shared between all three writer groups; red shading denotes families shared between L2S/L1S writers; yellow shading indicates items shared between L2S and Published writers; and blue shading shows those families shared between L1S and Published writers.

Table 48b show the extent of item sharing between writer groups in the form of percentages of the total usage within the top 20 items.

Table 48a-20 most frequently used AWL families for each writer group

L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	2074	3.25	research	4012	3.29	research	7002	2.54
participate	1219	5.15	analyse	2063	4.98	participate	5957	4.71
analyse	886	6.54	identify	2037	6.65	analyse	4217	6.24
strategy	851	7.87	individual	1887	8.19	respond	3658	7.57
communicate	827	9.17	participate	1550	9.46	consume	3596	8.87
culture	811	10.44	data	1522	10.71	significant	3455	10.13
data	804	11.7	focus	1510	11.95	data	3417	11.37
method	786	12.93	theory	1385	13.08	identify	3227	12.54
identify	782	14.15	create	1369	14.2	process	3081	13.66
focus	777	15.37	culture	1286	15.25	individual	2979	14.74
append	728	16.51	media	1219	16.25	focus	2924	15.8
motive	714	17.62	theme	1188	17.23	vary	2913	16.86
approach	684	18.7	approach	1176	18.19	perceive	2745	17.86
process	680	19.76	factor	1174	19.15	positive	2438	18.74
create	652	20.78	range	1103	20.06	context	2418	19.62
task	643	21.79	method	1072	20.93	strategy	2331	20.47
individual	643	22.79	consume	1067	21.81	approach	2276	21.29
attitude	563	23.67	evident	1053	22.67	assess	2237	22.11
context	553	24.54	strategy	1029	23.51	involve	2058	22.85
design	537	25.38	communicate	981	**24.32**	specific	2054	**23.6**

Table 48b-Top 20 AWL families item sharing between writer groups

	L2S	L1S	Published
Total frequency / % of total top 20 usage	16214 / 100	29683 / 100	64983 / 100
Shared between all groups frequency / % of total top 20 usage	8720 / 53.78	16768 / 56.55	34330 / 52.83
Shared between L2S/L1S frequency / % of total top 20 usage	3076 / 18.97	4708 / 15.86	n/a
Shared between L2S/Published frequency / % of total top 20 usage	1233 / 7.60	n/a	5499 / 8.46
Shared between L1S/Published frequency / % of total top 20 usage	n/a	1067 / 3.59	3596 / 5.53
Total shared % of total top 20 usage	80.36	76.01	66.83
Unshared frequency / % of total top 20 usage	3185 / 19.64	7122 / 23.99	21558 / 33.17

Nine top 20 AWL families were shared between all three writer groups, accounting for over

50% of top 20 usage for each group, and including the three most frequently used families for

Published writers, the four most frequently used families for L2S writers, and the seven most

frequently used families for L1S writers. If two and three-three way sharing are combined[4], the student groups both shared more similarity with each other than they did with the Published group. Total sharing[5] was highest in the L2S group and lowest in the Published group, for whom ~33% of top 20 usage represented items not found in the other two groups. It is interesting to note that while a number of families one might expect to occur commonly in academic writing are shared between all three group (*research, participate, analyse* and *data*, for example), *significant,* which is the 6th most frequently used family in the Published group, does not appear in the top 20 of either student writer group. This may reflect the more practical, research-oriented focus of research articles, and the associated presentation of data and statistical analysis, which would not be necessary in many kinds of student assignment.

4.4.1.2 Lexical Bundles

Table 49a shows the 20 most frequent 4-word bundles, with frequencies (F), for the three writer groups across the corpus as a whole. Bundle sharing between writer groups is denoted using the same colour coding system as used in the preceding section. Tables 49b shows the extent of item sharing between writer groups in the form of percentages of the total usage within the top 20 items.

---

[4] In the case of L2S for example, combining items shared between all three groups and items shared between L2S/L1S only, would provide a complete picture of what L2S writers shared with L1S writers overall
[5] For L2S for example – items shared between three groups, items shared between L2S/L1S, and items shared between L2S/Published.

Table 49a-20 most frequently used 4-word bundles for each writer group

L2S		L1S		Published	
Bundle	F	Bundle	F	Bundle	F
on the other hand	135	it is important to	180	in the context of	300
it is important to	91	as a result of	102	on the other hand	207
at the end of	75	as well as the	94	the extent to which	203
is one of the	66	to be able to	90	it is important to	193
one of the most	63	will be able to	89	at the same time	185
the end of the	62	the criminal justice system	85	are more likely to	157
when it comes to	57	are more likely to	85	as well as the	149
to be able to	56	when it comes to	83	as a result of	147
students with special needs	56	on the other hand	80	on the basis of	130
at the same time	56	the end of the	77	in the case of	129
will be able to	52	at the end of	77	in the united states	128
as well as the	47	in the form of	70	the purpose of this	125
in the target language	43	it could be argued	69	the end of the	119
as a result of	43	the way in which	68	in relation to the	112
in the field of	42	one of the most	68	in the form of	105
effective teaching and learning	41	the use of the	66	at the end of	105
the ministry of education	39	is one of the	66	the ways in which	103
the rest of the	36	in the case of	66	in the present study	103
to make sure that	35	the rest of the	61	in terms of the	101
can be seen in	34	in relation to the	59	more likely to be	95

Table 49b-Top 20 4-word bundles item sharing between writer groups

	L2S	L1S	Published
Total frequency / % of total top 20 usage	1129/ 100	1635/ 100	2896 / 100
Shared between all groups frequency / % of total top 20 usage	453 / 40.12	610 / 37.31	920 / 31.77
Shared between L2S/L1S frequency / % of total top 20 usage	330 / 29.23	457 / 27.95	n/a
Shared between L2S/Published frequency / % of total top 20 usage	0 / 0.00	n/a	0 / 0.00
Shared between L1S/Published frequency / % of total top 20 usage	n/a	155 / 9.48	262 / 9.05
Total shared % of total top 20 usage	69.35	74.74	40.81
Unshared frequency / % of total top 20 usage	346 / 30.65	413 / 25.26	1714 / 59.19

Six top 20 bundles were shared across all three writer groups, the same number as were

shared between L2S/L1S writers. With two and three-way sharing combined, both student

groups shared more with each other than they did with the published group. Total sharing

was also lowest among published writers, with ~60% of top 20 usage being unique to that

group (as opposed to ~30% for L2S writers, and ~25% for L1S writers). Some unshared items in the L2S and L1S groups(*students with special needs, in the target language, the criminal justice system, effective teaching and learning* and the *Ministry of Education*) would seem to have made it into the top 20 as a result of high occurrence in specific disciplines, rather than through being widely distributed across the whole corpus.

4.4.1.3 Hedging

Table 50a shows the 10 most frequent lexical hedges, with frequencies (F) and percentage of total hedge usage for the three writer groups within the corpus as a whole. Sharing is indicated using the same system outlined previously. Table 50b shows the extent of item sharing between writer groups in the form of percentages of the total usage within the top 10 items

Table 50a-10 most frequently used lexical hedges for each writer group

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
could	679	11.95	may	2353	19.04	may	4079	17.18
might	666	11.72	suggest-	1599	12.94	suggest-	2191	9.23
may	651	11.46	could	1564	12.66	indicate-	1560	6.57
suggest-	399	7.02	often	763	6.18	could	1446	6.09
seem-	287	5.05	likely	720	5.83	likely	1342	5.65
often	262	4.61	argue-	450	3.64	might	1320	5.56
argue-	227	3.99	seem-	402	3.25	often	1278	5.38
indicate-	170	2.99	indicate	390	3.16	seem-	743	3.13
usually	159	2.80	feel-	389	3.15	argue-	733	3.09
likely	159	2.80	appear-	338	2.74	appear-	588	2.48
Total	3659	64.39	Total	8968	72.59	Total	15280	64.36

	L2S	L1S	Published
Total frequency / % of total top 10 usage	3659 / 100	8968 / 100	15280 / 100
Shared between all groups frequency / % of total top 10 usage	2834 /77.45	8241 /91.89	13372 / 87.51
Shared between L2S/L1S frequency / % of total top 10 usage	0 / 0.00	0 /0.00	n/a
Shared between L2S/Published frequency / % of total top 10 usage	666 / 18.20	n/a	1320 / 8.64
Shared between L1S/Published frequency / % of total top 10 usage	n/a	338 /.3.77	588 / 3.85
Total shared % of total top 10 usage	95.65	95.66	100
Unshared frequency / % of total top 10 usage	159 /4.35	389 /4.34	0 /0.00

Top ten items made up a similar proportion of total usage in the L2S and Published groups (~64%), although this was slightly higher among L1S writers, at 72.59%. Eight of the top 10 items were common to all three writer groups, suggesting a high degree of similarity between them in terms of the most commonly used lexical hedges. *May* was the most frequently used lexical hedge for both the L1S and Published writers. With two and three-way sharing combined, both student groups bore more similarity to the Published writers than they did to each other. Total sharing in all writer groups was above 95% of top 10 usage, with Published writers having no unshared items in the top 10.

4.4.1.4 Citations

Table 51 shows rates of overall citation per 1000 words, as well as use of integral/non-integral citations and the various integral forms, expressed as percentages of total integral citation use for the three writer groups across the whole corpus.

Table 51-Overall citation use for each writer group

	L2S	L1S	Published
Citations/1000 words	10.11	9.41	6.71
Non-integral citations %	66.32	62.26	69.21
Integral citations %	33.68	37.74	30.79
Non-Reporting integral citations %	17.35	14.37	23.67
According to %	14.23	4.02	4.22
Discourse act verbs %	45.63	47.46	38.06
Research act verbs %	17.59	27.36	29.57
Cognition act verbs %	5.11	6.57	3.64
Ambiguous %	0.09	0.22	0.84

Published writers showed the lowest overall citation rate and the highest relative proportion of non-integral citations. Non-reporting structures were lowest among the L1S group and highest among Published writers. Proportionally, *according to* was used with notably more frequency by L2S writers than by either of the other groups. In terms of reporting verb categories, discourse and cognition act verbs were proportionally lowest among Published writers, while the L2S group showed the lowest proportion of research act verbs. In all three groups, discourse act verbs were the most common, and cognition acts used least often.

Table 52a shows the 10 most frequent reporting verbs, with frequencies (F) and percentage of total integral citation use for the three writer groups across the corpus as a whole. Sharing is indicated as in preceding sections. Table 52b shows the extent of item sharing between writer groups in the form of percentages of the total usage within the top 10 items.

Table 52a-10 most frequently used reporting verbs for each writer group

L2S			L1S			Pub		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	301	14.23	suggest	398	9.83	find	319	6.96
state	120	5.67	find	375	9.26	suggest	227	5.03
suggest	119	5.63	state	308	7.60	argue	189	4.28
argue	90	4.26	argue	165	4.07	according to	188	4.22
find	80	3.78	according to	163	4.02	describe	135	3.02
mention	59	2.79	describe	111	2.74	report	131	2.83
claim	55	2.60	note	95	2.35	show	117	2.61
point out	39	1.84	conduct	88	2.17	note	112	2.47
define	33	1.56	discuss	71	1.75	identify	96	2.15
describe	32	1.51	propose	70	1.73	propose	88	1.97
show	32	1.51	support	70	1.73			
Total %		45.38	Total %		47.25	Total %		35.54

Table 52b-Top 10 reporting verbs item sharing between writer groups

	L2S	L1S	Published
Total frequency / % of total top 10 usage	960/100	1914/100	1602/100
Shared between all groups frequency / % of total top 10 usage	622/64.79	1212/63.32	1058/66.04
Shared between L2S/L1S frequency / % of total top 10 usage	120/12.50	308/16.09	n/a
Shared between L2S/Published frequency / % of total top 10 usage	32/3.33	n/a	117/7.30
Shared between L1S/Published frequency / % of total top 10 usage	n/a	165/8.62	200/12.48
Total shared % of total top 10 usage	80.63	88.04	85.83
Unshared frequency / % of total top 10 usage	186/19.37	229/11.96	227/14.17

The proportion of total integral citation usage represented by the ten most frequently used items was highest among L1S writers (47.25%) and lowest among Published writers (35.54%), suggesting that the latter writers rely less-heavily than their student counterparts on the most common reporting verbs. Five items were shared between all three groups, representing over 63% of top 10 usage in each group. With two and three-way sharing combined, the student groups both bore more similarity to each other than they did to the published group. Total sharing was above 80% in all groups, but was highest among L1S writers, at 88.04%, and lowest among L2S writers, at 80.63%.

4.4.2 Within disciplines

4.4.2.1 Vocabulary

This analysis compared vocabulary usage among the three writer groups within each discipline. As per the process described in Section 4.3.1, the L2S, L1S and Published sub-corpora were normalised and balanced for the smallest of the three within each discipline so as to facilitate direct comparison of AWL and GSL % type coverage in each writer group; details of the normalisation are given in Appendix C.

4.4.2.1(i) AWL and GSL Coverage

Table 53 shows AWL and GSL 1K and 2K coverage for the three writer groups within each discipline, along with Off-List type coverage and AWL families.

Table 53-AWL and GSL coverage for writer groups within each discipline

	Advertising			Education		
	L2S	L1S	Pub	L2S	L1S	Pub
1K and 2K type %	46.80	46.70	41.82	46.37	50.87	43.33
AWL type %	17.46	16.42	20.17	21.15	19.78	20.96
Off-List type %	35.74	36.89	38.01	32.48	29.35	35.70
AWL families	428	353	432	423	393	448
	Fashion			Forensic Psychology		
	L2S	L1S	Pub	L2S	L1S	Pub
1K and 2K type %	44.77	44.47	41.85	46.78	46.27	43.50
AWL type %	16.73	16.46	19.98	20.86	19.78	20.27
Off-List type %	38.50	39.07	38.17	32.36	33.95	36.23
AWL families	429	410	461	401	380	410
	Languages			Physiotherapy		
	L2S	L1S	Pub	L2S	L1S	Pub
1K and 2K type %	46.12	44.20	40.31	57.41	53.66	53.88
AWL type %	18.72	18.71	19.77	22.75	20.75	21.44
Off-List type %	33.15	37.10	39.92	19.84	25.58	24.68
AWL families	442	462	482	220	203	188
	Social Work			TESOL		
	L2S	L1S	Pub	L2S	L1S	Pub
1K and 2K type %	50.18	52.67	49.08	33.63	32.25	29.55
AWL type %	22.31	21.41	21.68	16.09	16.07	14.96
Off-List type %	27.51	25.93	29.24	50.27	51.68	55.49
AWL families	325	300	341	537	552	547

In seven of the eight disciplines, L2S writers had the highest level of 1K and 2K coverage, the exception being Social Work, in which 1K and 2K coverage was highest among L1S writers. 1K and 2K coverage was lowest among Published writers in every discipline except Physiotherapy; in this discipline L1S 1K and 2K coverage was marginally lower than that of Published writers. AWL coverage was highest among Published writers in Advertising, Fashion, and Languages, but was highest among L2S writers in all other disciplines. While the lowest level of AWL coverage in TESOL was shown by Published writers, in the other seven disciplines it was L1S writers who demonstrated the lowest AWL type coverage. It may be that L2S writers, given the additional language pressures they often face in English academic writing, focus more acutely on their use of vocabulary, on trying to write 'academically', and on attempting to use

a variety of academic vocabulary, which may explain their generally higher levels of AWL coverage when compared to the L1S writers.

In six of the eight disciplines, Off-list coverage was highest among Published writers, with L1S writers showing the highest level only in Fashion and Physiotherapy. Lowest levels of Off-List coverage were found among L2S writers in Advertising, Forensic Psychology, Languages, Physiotherapy, and TESOL; among L1S writers in Education and Social Work; and among Published writers in Fashion. In terms of AWL families, five of the disciplines (Advertising, Education, Fashion, Forensic Psychology, and Social Work) showed a similar pattern, with the highest number of AWL families shown by Published writers, and the lowest by L1S writers. Disciplines that differed were Languages (L2S lowest), Physiotherapy (Published lowest, L2S highest), and TESOL (L2S lowest, L1S highest).

4.4.2.1(ii) Frequency of AWL Vocabulary

To investigate AWL vocabulary frequency, all sub-corpora were analysed in their entirety. Tables 54a-61a show the 20 most frequent AWL families, with frequencies and cumulative percentages for the three writer groups within each of the disciplines in turn. They also indicate the extent of item sharing. Family sharing between writer groups is denoted in the following way: White shading indicates families shared between all three writer groups; red shading denotes families shared between L2S/L1S writers; yellow shading indicates items shared between L2S and Published writers; and blue shading shows those families shared between L1S and Published writers.

Tables 54b-61b show the extent of item sharing between writer groups in the form of percentages of the total usage within the top 20 items.

4.4.2.1(ii)a Advertising

Table 54a-20 most frequently used AWL families for each writer group in Advertising

Advertising								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	159	3.31	research	543	4.12	consume	2138	5.05
create	136	6.14	create	374	6.96	research	1159	7.79
consume	103	8.29	gender	325	9.43	perceive	922	9.97
philosophy	99	10.35	target	277	11.53	respond	882	12.06
culture	98	12.39	analyse	268	13.57	media	840	14.04
ethic	76	13.97	consume	268	15.60	purchase	781	15.89
identify	73	15.49	media	267	17.63	attitude	773	17.72
image	72	16.99	ethic	219	19.29	positive	743	19.47
focus	71	18.47	identify	213	20.91	individual	731	21.20
gender	69	19.90	image	200	22.43	significant	671	22.79
individual	67	21.30	culture	192	23.88	identify	605	24.22
media	63	22.61	generation	181	25.26	vary	579	25.58
minor	58	23.82	focus	170	26.55	process	536	26.85
ethnic	56	24.98	role	163	27.79	analyse	535	28.12
target	54	26.11	communicate	139	28.84	create	489	29.27
communicate	51	27.17	similar	134	29.86	participate	478	30.40
role	50	28.21	feature	131	30.85	communicate	467	31.51
impact	48	29.21	previous	126	31.81	strategy	445	32.56
specific	46	30.17	range	118	32.71	data	417	33.54
perceive	46	31.13	brief	116	33.59	environment	410	34.51

Table 54b-Top 20 AWL families item sharing between writer groups in Advertising

	L2S	L1S	Published
Total frequency / % of total top 20 usage	1495 / 100	4424/ 100	14601 / 100
Shared between all groups frequency / % of total top 20 usage	585 / 39.13	1804 / 40.78	5698 / 39.02
Shared between L2S/L1S frequency / % of total top 20 usage	490 / 32.78	1546 / 34.95	n/a
Shared between L2S/Published frequency / % of total top 20 usage	113 / 7.56	n/a	1653 / 11.32
Shared between L1S/Published frequency / % of total top 20 usage	n/a	268 / 6.06	535 / 3.66
Total shared % of total top 20 usage	79.47	81.79	54.00
Unshared frequency / % of total top 20 usage	307 / 20.53	806 / 18.21	6715 / 46.00

Six top 20 families (*research, create, consume, identify, media* and *communicate*) were shared
between all three writer groups, representing ~40% of top 20 usage for each group, with
sharing between L2S/L1S writers constituting the next biggest proportion. If two-way and
three-way sharing are combined, the student groups both bore more similarity to each other

than they did the Published group. Total sharing was at a similar level for L2S and L1S writers, but was lower for Published writers, who showed a much higher proportion of unshared items unique to that writer group. Published writers in this discipline showed the highest total AWL families and the highest overall AWL coverage. It is possible that these writers make more frequent use of a wider range of AWL types and thus show less commonality with the student groups.

4.4.2.1(ii)b Education

Table 55a-20 most frequently used AWL families for each writer group in Education

Education								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	1010	10.40	research	370	9.19	research	865	2.74
method	208	12.54	participate	105	11.80	participate	743	5.10
participate	199	14.58	method	78	13.74	data	563	6.88
data	181	16.45	achieve	73	15.55	analyse	539	8.59
individual	139	17.88	focus	68	17.24	focus	488	10.14
focus	129	19.21	approach	63	18.81	culture	424	11.48
approach	128	20.52	create	61	20.32	identify	400	12.75
process	125	21.81	data	59	21.79	significant	346	13.85
institute	125	23.10	involve	55	23.16	context	324	14.88
policy	124	24.37	access	54	24.50	design	324	15.90
communicate	122	25.63	ensure	54	25.84	individual	311	16.89
implement	120	26.86	individual	54	27.18	approach	309	17.87
achieve	114	28.04	benefit	48	28.37	process	300	18.82
ethic	112	29.19	environment	47	29.54	method	295	19.75
analyse	109	30.31	issue	44	30.63	perceive	294	20.69
require	97	31.31	process	41	31.65	community	289	21.60
issue	96	32.30	specific	40	32.65	project	288	22.52
gender	87	33.19	paradigm	39	33.61	respond	274	23.38
topic	86	34.08	require	38	34.56	involve	263	24.22
define	85	34.95	ethic	36	35.45	vary	260	25.04

Table 55b-Top 20 AWL families item sharing between writer groups in Education

	L2S	L1S	Published
Total frequency / % of total top 20 usage	3396 / 100	1427/ 100	7899 / 100
Shared between all groups frequency / % of total top 20 usage	2119 / 62.40	838 / 58.72	3874 / 49.04
Shared between L2S/L1S frequency / % of total top 20 usage	419/ 12.34	191 / 13.38	n/a
Shared between L2S/Published frequency / % of total top 20 usage	109 / 3.21	n/a	539 / 6.82
Shared between L1S/Published frequency / % of total top 20 usage	n/a	0 / 0.00	0 / 0.00
Total shared % of total top 20 usage	77.95	72.10	55.86
Unshared frequency / % of total top 20 usage	749/ 22.05	398 / 27.90	3486 / 44.14

Eight top 20 items (*research, method, participate, data, individual, focus, approach* and *process*) were shared between all three groups, representing around 60% of top 20 usage for L2S and L1S writers, and around 50% for Published writers. Item sharing between only two groups was comparatively low, but again, was higher between L2S/L1S. Overall, both student groups shared more with each other than they did with the Published writers. Total sharing was lowest among Published writers, for whom just under 45% of top 20 usage was made up of items not found in the other two groups. AWL type coverage among writer groups in this discipline was fairly uniform, while Published writers employed the highest number of AWL families; this may have resulted in fewer shared items for this group.

4.4.2.1(ii)c Fashion

Table 56a-20 most frequently used AWL families for each writer group in Fashion

Fashion								
L2S			L1S			Punlished		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
range	229	4.68	range	730	4.98	consume	1394	3.45
design	210	8.98	consume	603	9.10	research	925	5.75
consume	193	12.92	create	315	11.25	analyse	622	7.29
strategy	144	15.87	strategy	282	13.17	strategy	540	8.63
create	113	18.18	design	272	15.03	perceive	510	9.89
sustain	97	20.16	trend	245	16.70	process	465	11.04
finance	74	21.68	finance	220	18.20	invest	461	12.18
analyse	73	23.17	style	219	19.69	positive	452	13.30
label	70	24.60	technology	198	21.05	vary	451	14.42
style	67	25.97	focus	174	22.23	focus	446	15.53
ratio	66	27.32	sustain	172	23.41	respond	446	16.63
margin	64	28.63	analyse	167	24.55	corporate	432	17.70
trend	64	29.94	target	166	25.68	institute	422	18.75
culture	56	31.08	margin	153	26.72	create	421	19.79
append	53	32.17	media	148	27.73	innovate	416	20.82
item	49	33.17	invest	147	28.74	identify	409	21.84
process	48	34.15	item	142	29.71	theory	401	22.83
tradition	45	35.07	ratio	136	30.63	resource	401	23.82
media	43	35.95	purchase	136	31.56	significant	385	24.78
target	42	36.81%	indicate	125	32.41%	data	381	25.72%

Table 56b-Top 20 AWL families item sharing between writer groups in Fashion

	L2S	L1S	Published
Total frequency / % of total top 20 usage	1800 / 100	4750 / 100	10380 / 100
Shared between all groups frequency / % of total top 20 usage	523 / 29.06	1367 / 28.78	2977 / 28.68
Shared between L2S/L1S frequency / % of total top 20 usage	1005/ 55.83	2601 / 54.80	n/a
Shared between L2S/Published frequency / % of total top 20 usage	48 / 2.67	n/a	456 / 4.48
Shared between L1S/Published frequency / % of total top 20 usage	n/a	147 / 3.09	461 / 4.44
Total shared % of total top 20 usage	87.56	86.67	37.60
Unshared frequency / % of total top 20 usage	224 / 12.44	633 / 13.33	6477 / 62.40

Around 29% of top 20 usage in each writer group could be attributed to the four families shared between all three groups (*consume, strategy, create* and *analyse*). Sharing amongst only the two student writer groups constituted a much larger proportion of their usage in

comparison (L2S – 55,83%, L1S – 54.80%), while sharing between L2S/Published only or L1S/Published only was considerably lower, indicating that in this discipline too, student groups have more in common with each other than they do with published writers. This is confirmed when two and three-way sharing are combined: L2S and L1S groups both shared around 84% of their top 20 usage with each other, compared to the roughly 33% shared by the Published writers with either student group. Overall sharing for L2S and L1S groups was over 85%, compared to 37.6% for Published writers. This latter group again showed the highest proportion of unshared items from any of the three groups, at 63.40% of top 20 usage, perhaps attributable to higher AWL type coverage, and higher total AWL families.

4.4.2.1(ii)d Forensic Psychology

Table 57a-20 most frequently used AWL families for each writer group in Forensic Psychology

Forensic Psychology								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
research	213	4.62	research	1908	5.78	participate	912	3.33
individual	118	7.18	identify	958	8.69	research	639	5.67
identify	98	9.30	theme	896	11.40	significant	565	7.73
factor	95	11.36	individual	895	14.11	individual	451	9.38
media	86	13.23	analyse	758	16.41	analyse	448	11.01
analyse	73	14.81	evident	681	18.48	factor	377	12.39
psychology	65	16.22	media	666	20.49	approach	365	13.73
evident	63	17.59	data	618	22.37	respond	362	15.05
mental	58	18.85	factor	580	24.13	mental	361	16.37
data	57	20.08	mental	441	25.46	vary	355	17.66
domestic	57	21.32	sex	435	26.78	identify	348	18.93
community	56	22.53	participate	354	27.85	incidence	318	20.10
technique	56	23.75	previous	326	28.84	assess	315	21.25
culture	51	24.85	partner	315	29.80	data	299	22.34
investigate	51	25.96	negate	315	30.75	strategy	276	23.35
positive	50	27.04	issue	291	31.63	consist	270	24.33
approach	50	28.13	physical	280	32.48	indicate	261	25.29
theme	47	29.15	proceed	275	33.32	sex	257	26.23
perceive	47	30.17	conduct	272	34.14	specific	256	27.16
respond	46	31.16	psychology	264	34.94	process	253	28.09

Table 57b-Top 20 AWL families item sharing between writer groups in Forensic Psychology

	L2S	L1S	Published
Total frequency / % of total top 20 usage	1437 / 100	11528/ 100	7688 / 100
Shared between all groups frequency / % of total top 20 usage	712 / 49.55	6158 / 53.42	2923 / 38.02
Shared between L2S/L1S frequency / % of total top 20 usage	261 / 18.16	2507 / 21.75	n/a
Shared between L2S/Published frequency / % of total top 20 usage	96 / 6.68	n/a	727 / 9.46
Shared between L1S/Published frequency / % of total top 20 usage	n/a	789 / 6.84	1169 / 15.21
Total shared % of total top 20 usage	74.39	82.01	62.69
Unshared frequency / % of total top 20 usage	368 / 25.61	2074 / 17.99	2869 / 37.31

Sharing between all three groups constituted roughly 50% of top 20 usage for L2S and L1S writers, and just under 39% for Published writers. This translates to seven families (*research, individual, identify, factor, analyse, mental* and *data*) and was the largest proportion of usage for all three writer groups. For L1S writers, the next largest proportion was sharing with L2S students, while for L2S and Published writers, unshared items made up the second largest proportion of top 20 usage. Overall, student groups shared more with each other than they did with the Published writers. As in the preceding disciplines, Published writers showed the highest proportion of unshared items, and correspondingly, the lowest proportion of overall sharing. Highest total sharing was shown in this discipline by L1S writers (82.01%), with L2S writers slightly lower at 74.39%.

4.4.2.1(ii)e Languages

Table 58a-20 most frequently used AWL families for each writer group in Languages

Languages								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
strategy	208	2.52	target	189	2.61	research	1075	3.13
target	191	4.84	task	178	5.06	participate	856	5.63
task	186	7.10	append	169	7.39	process	617	7.43
append	154	8.97	focus	150	9.46	assess	580	9.12
focus	144	10.72	strategy	136	11.34	policy	544	10.70
create	138	12.39	challenge	119	12.98	context	529	12.24
instruct	133	14.01	differentiate	109	14.48	analyse	472	13.62
environment	119	15.45	create	102	15.89	approach	460	14.96
achieve	119	16.90	achieve	100	17.27	focus	452	16.28
motive	118	18.33	communicate	100	18.65	professional	429	17.53
challenge	111	19.67	ensure	94	19.94	data	414	18.73
communicate	108	20.99	environment	88	21.16	task	371	19.81
positive	104	22.25	motive	88	22.37	identify	357	20.85
factor	103	23.50	approach	85	23.54	evaluate	324	21.80
process	101	24.72	instruct	81	24.66	text	318	22.72
culture	93	25.85	individual	72	25.66	specific	311	23.63
research	92	26.97	require	72	26.65	academy	307	24.53
approach	90	28.06	academy	72	27.64	individual	303	25.41
individual	81	29.04	factor	71	28.62	respond	301	26.29
topic	80	30.02	assess	69	29.57	strategy	299	27.16

Table 58b-Top 20 AWL families item sharing between writer groups in Languages

	L2S	L1S	Published
Total frequency / % of total top 20 usage	2473 / 100	2144 / 100	9319 / 100
Shared between all groups frequency / % of total top 20 usage	709 / 28.67	621 / 28.96	1885 / 20.23
Shared between L2S/L1S frequency / % of total top 20 usage	1294 / 52.33	1107 / 51.63	n/a
Shared between L2S/Published frequency / % of total top 20 usage	193 / 7.80	n/a	16.92 / 18.16
Shared between L1S/Published frequency / % of total top 20 usage	n/a	141 / 6.58	887 / 9.52
Total shared % of total top 20 usage	88.80	87.17	47.91
Unshared frequency / % of total top 20 usage	277 / 11.20	275 / 12.83	4855 / 52.09

Five top 20 AWL families (*strategy, task, focus, approach* and *individual*) were shared between

all three writer groups, constituting ~29% of top 20 usage for L2S and L1S students, and just

over 20% for Published writers. L2S/L1S sharing was by far the largest proportion of usage for

these two groups, and with two and three-way sharing combined, it is clear that in this discipline too, student writers showed more commonality each other than either of them did with the Published group. In terms of total sharing, levels among L2S and L1S students were similar (88.80% and 87.17% respectively). For Published writers, this was considerably lower, at 47.91%. Unshared items amounted to 52.09% (11 items) for Published writers, who also showed the highest AWL type coverage and highest total AWL families in this discipline.

4.4.2.1(ii)f Physiotherapy

Table 59a-20 most frequently used AWL families for each writer group in Physiotherapy

Physiotherapy								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
team	67	8.38	team	95	3.64	participate	866	4.06
culture	33	12.50	culture	73	6.44	injure	744	7.55
role	26	15.75	individual	66	8.98	physical	674	10.72
task	18	18.00	evident	66	11.51	data	450	12.83
positive	17	20.13	injure	53	13.54	intervene	417	14.79
distribute	16	22.13	role	51	15.50	analyse	413	16.72
implement	15	24.00	approach	41	17.07	assess	411	18.65
individual	14	25.75	professional	39	18.57	factor	368	20.38
challenge	13	27.38	challenge	36	19.95	significant	324	21.90
structure	12	28.88	style	35	21.29	outcome	317	23.39
achieve	11	30.25	focus	35	22.63	vary	303	24.81
establish	11	31.63	outcome	35	23.97	medical	277	26.11
analyse	10	32.88	significant	32	25.20	function	275	27.40
create	10	34.13	assess	32	26.43	research	233	28.49
vision	9	35.25	promote	30	27.58	specific	217	29.51
goal	9	36.38	identify	30	28.73	evaluate	212	30.50
involve	9	37.50	demonstrate	28	29.80	team	198	31.43
outcome	9	38.63	research	26	30.80	individual	196	32.35
professional	8	39.63	potential	26	31.80	respond	195	33.27
strategy	8	40.63	strategy	26	32.80	previous	189	34.15

Table 59b-Top 20 AWL families item sharing between writer groups in Physiotherapy

	L2S	L1S	Published
Total frequency / % of total top 20 usage	325 / 100	855 / 100	7279 / 100
Shared between all groups frequency / % of total top 20 usage	90 / 27.69	196 / 22.92	711 / 9.77
Shared between L2S/L1S frequency / % of total top 20 usage	88 / 27.08	225 / 26.32	n/a
Shared between L2S/Published frequency / % of total top 20 usage	10 / 3.08	n/a	413 / 5.67
Shared between L1S/Published frequency / % of total top 20 usage	n/a	143 / 16.73	1712 / 23.52
Total shared % of total top 20 usage	57.85	65.97	38.96
Unshared frequency / % of total top 20 usage	137 / 42.15	291 / 34.03	4443 / 61.04

Only three top 20 families (*team, individual* and *outcome*) were shared between all three writer groups, and constituted 27.69% of top 20 usage for L2S writers, 22.92% for L1S writers, and 9.77% for Published writers. Once again, if sharing between two groups is combined with sharing between three, both student groups can be seen to share more with each other than they did with the Published group. For L2S and L1S writers, proportions of usage attributable to unshared items were relatively high compared to other disciplines, at 42.15% and 34.03% respectively. Published writers showed the highest proportion of unshared items (61.04%), although in this discipline neither the highest AWL type coverage, nor the highest total AWL families, so perhaps in this case the high level of unshared items is simply due the presence of a few, very frequently used families (notably - *participate, data, intervene*) that did not occur within the top 20 usage of either student group. This could reflect a difference in text type or purpose; given that families such as *participate, data,* and *intervene* would be expected to appear with a reasonably high frequency in research articles based on practical research, it may be that the student groups' usage differed as a result of different tasks, perhaps more theoretical in nature.

4.4.2.1(ii)g Social Work

Table 60a-20 most frequently used AWL families for each writer group in Social Work

Social Work								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
theory	176	7.68	theory	798	9.21	research	1100	2.54
focus	75	10.95	approach	341	13.15	participate	790	4.37
approach	62	13.65	individual	292	16.52	community	561	5.66
individual	53	15.96	focus	210	18.95	culture	543	6.92
intervene	45	17.92	intervene	198	21.24	respond	543	8.17
issue	43	19.80	issue	175	23.26	identify	432	9.17
assess	42	21.63	impact	129	24.75	theory	420	10.14
adult	38	23.29	attach	127	26.21	individual	419	11.11
identify	36	24.86	domestic	118	27.58	positive	403	12.04
furthermore	35	26.38	method	118	28.94	analyse	379	12.92
psychology	34	27.87	identify	117	30.29	significant	376	13.79
professional	30	29.18	positive	110	31.56	focus	363	14.63
attach	29	30.44	role	104	32.76	sex	362	15.46
impact	28	31.66	recover	97	33.88	process	345	16.26
process	23	32.66	mental	94	34.97	data	338	17.04
aspect	23	33.67	adult	91	36.02	vary	325	17.79
occur	23	34.67	professional	85	37.00	professional	321	18.53
partner	21	35.59	demonstrate	79	37.91	intervene	313	19.26
analyse	21	36.50	perspective	77	38.80	factor	304	19.96
environment	20	37.37	assess	73	39.64	involve	304	20.66

Table 60b-Top 20 AWL families item sharing between writer groups in Social Work

	L2S	L1S	Published
Total frequency / % of total top 20 usage	857 / 100	3433 / 100	8941 / 100
Shared between all groups frequency / % of total top 20 usage	415 / 48.42	1700 / 49.52	2268 / 25.37
Shared between L2S/L1S frequency / % of total top 20 usage	242 / 28.24	936 / 27.26	n/a
Shared between L2S/Published frequency / % of total top 20 usage	21 / 2.45	n/a	379 / 4.24
Shared between L1S/Published frequency / % of total top 20 usage	n/a	110 / 3.20	403 / 4.51
Total shared % of total top 20 usage	79.11	79.98	34.12
Unshared frequency / % of total top 20 usage	179 / 20.89	687 / 20.02	5891 / 65.88

Six top 20 families (*theory, focus, individual, intervene, identify* and *professional*) were shared

between all three groups, constituting almost 50% of usage for L2S and L1S writers, and just

over 25% for Published writers, suggesting that these few common families are perhaps relied

upon to a greater extent by the student writers. AWL families shared between L2S/L1S writers

amounted to 28.24% for L2S writers, and 27.26% for L1S writers. In comparison, both groups

of student writers showed much lower proportions when it came to items shared only with

Published writers, so as in other disciplines, when overall sharing between the different groups

is considered, the primary commonality is between the student groups. Total item sharing for

L2S and L1S writers made up almost 80% of top 20 usage, compared to just under 35% for

Published writers, who again showed the highest total AWL families within this discipline.

Unshared items in the Published group included families such as *research, participate, data*,

and *significant*, which may again point to usage variation arising from the differences between

practical research articles and student assignments.

4.4.2.1(ii)h TESOL

Table 61a-20 most frequently used AWL families for each writer group in TESOL

TESOL								
L2S			L1S			Published		
Family	Freq	cum %	Family	Freq	cum %	Family	Freq	cum %
participate	886	3.11%	participate	958	2.48%	task	1106	2.52
research	563	5.08%	research	927	4.87%	participate	1029	4.86
append	518	6.89%	analyse	764	6.84%	research	1006	7.16
analyse	517	8.71%	culture	756	8.80%	analyse	809	9.00
motive	512	10.50	data	675	10.54	context	679	10.55
data	495	12.24	motive	601	12.09	respond	655	12.04
communicate	489	13.95	identify	538	13.48	instruct	617	13.44
identify	438	15.48	append	486	14.74	focus	571	14.74
method	429	16.99	assess	471	15.96	interact	558	16.02
strategy	418	18.45	focus	453	17.13	data	555	17.28
task	400	19.86	task	446	18.28	text	532	18.49
culture	400	21.26	method	433	19.40	significant	499	19.63
attitude	389	22.62	respond	430	20.51	identify	487	20.74
context	370	23.92	communicate	423	21.60	process	481	21.83
process	310	25.00	strategy	396	22.63	communicate	470	22.91
approach	306	26.08	attitude	395	23.65	vary	434	23.89
evaluate	290	27.09	specific	392	24.66	construct	376	24.75
focus	268	28.03	approach	383	25.65	item	368	25.59
assess	245	28.89	context	371	26.61	create	358	26.41
status	122	29.32	individual	355	27.52	motive	337	27.17

Table 61b-Top 20 AWL families item sharing between writer groups in TESOL

	L2S	L1S	Published
Total frequency / % of total top 20 usage	8365 / 100	10653 / 100	11927 / 100
Shared between all groups frequency / % of total top 20 usage	4938 / 59.03	6156 / 57.79	7049/ 59.10
Shared between L2S/L1S frequency / % of total top 20 usage	2705 / 32.34	3320 / 31.16	n/a
Shared between L2S/Published frequency / % of total top 20 usage	310 / 3.71	n/a	481 / 4.03
Shared between L1S/Published frequency / % of total top 20 usage	n/a	430 / 4.04	655 / 5.49
Total shared % of total top 20 usage	95.08	92.99	68.62
Unshared frequency / % of total top 20 usage	412 / 4.92	747 / 7.01	3742 / 31.38

Ten of the top 20 AWL families were shared between all three writer groups, constituting almost 60% of usage in each of the groups. With two and three-way sharing considered together, commonality is most pronounced between the two student groups. Total sharing

was higher in this discipline than in others for all three writer groups. Correspondingly, the highest level of unshared items was, as in other disciplines, found among Published writers. In TESOL the Published group showed neither the highest total AWL families, nor the highest AWL coverage (in fact they were lowest of the three), and the unshared items do not suggest a particular RA vs. assignment distinction. It may simply be therefore that there is greater variety in the most commonly used AWL families within the Published group when compared to L2S/L1S writers, who both rely more heavily on a similar, limited selection of families.

4.4.2.1(iii) Summary of Within-Disciplines Results

While the various disciplines exhibited differing levels of within-discipline commonality between writer groups, there were patterns to be observed. L2S and L1S writers showed a clear tendency towards greater commonality with each other than with the Published writers, although whether this is due to differences between research articles and university assignments, or differences in knowledge and awareness of academic vocabulary is unclear. Similarly, Published writers in all eight disciplines exhibited lower levels of items sharing than the student groups. There may be a number of reasons for this, including a greater variety within the most commonly used AWL families compared to the student writers; effects of topic, and differences between the lexical demands of research articles and university assignments.

 Looking at the most commonly shared AWL families themselves within each discipline, there is some evidence of disciplinary variation. Items shared between all three groups include *consume, media* and *communicate* in Advertising; *create* and *strategy* in Fashion; *mental* in Forensic Psychology; *task* and *approach* in Languages; *team* and *outcome* in Physiotherapy; *intervene* and *professional* in Social Work, and *motive* and *communicate* in TESOL. While these

may not represent items specific to a given discipline, they may be an indication of specific lexical items being of increased relevance in some academic fields compared to others.

4.4.2.2 Lexical Bundle Use Within Disciplines

Tables 62a-69a show the 20 most frequent 4-word bundles, with frequencies (F), for the three writer groups within each of the disciplines in turn. Bundle sharing between writer groups is denoted as follows: White shading - bundles shared between all three writer groups; red shading - bundles shared between L2S/L1S writers; yellow shading - items shared between L2S and Published writers; and blue shading - bundles shared between L1S and Published writers. Tables 62b-69b show the extent of item sharing between writer groups in the form of percentages of the total usage within the top 20 items.

4.4.2.2(i) Advertising

Table 62a-20 most frequently used 4-word bundles for each writer group in Advertising

L2S	F	L1S	F	Published	F
it is important to	11	the representation of women	31	in the context of	71
on the other hand	10	representation of women in	30	are more likely to	47
one of the most	8	of the male body	16	on the other hand	40
are more likely to	8	it is important to	16	the extent to which	34
the last couple of	6	on the other hand	15	attitude toward the brand	31
in the last couple	6	the tone of voice	14	in the case of	29
gaps have been identified	6	the gender of the	12	one of the most	28
the way in which	5	of women in advertising	12	the main effect of	24
of the most important	5	as well as the	12	attitude toward the ad	22
more likely to be	5	a wide range of	12	the purpose of this	21
a critical evaluation of	5	portrayal of women in	11	is one of the	21
when it comes to	4	on the representation of	11	at the same time	21
to be able to	4	when it comes to	10	the relationship between the	20
the way people think	4	to the target audience	10	as a result of	20
the research question is	4	one of the most	10	a higher level of	18
part of a community	4	John Lewis is a	10	the results of the	17
of the brand and	4	in the United States	10	as well as the	17
last couple of years	4	will be able to	9	the mediating role of	16
it is also a	4	that there is a	9	respondents were asked to	16
is part of the	4	of gender in advertising	9	is positively related to	16

Table 62b-Top 20 4-word bundles item sharing between writer groups in Advertising

	L2S	L1S	Published
Total frequency / % of total top 20 usage	111 / 100	269 / 100	529 / 100
Shared between all groups frequency / % of total top 20 usage	18 / 16.22	25 / 9.29	68 / 12.85
Shared between L2S/L1S frequency / % of total top 20 usage	15 / 13.51	26 / 9.67	n/a
Shared between L2S/Published frequency / % of total top 20 usage	8 / 7.21	n/a	47 / 8.89
Shared between L1S/Published frequency / % of total top 20 usage	n/a	12 / 4.46	17 / 3.21
Total shared % of total top 20 usage	36.94	23.42	24.95
Unshared frequency / % of total top 20 usage	70 / 64.06	206 / 76.58	397 / 75.05

Sharing between writer groups was quite limited. Two top 20 items were shared between all three groups, two between L2S/L1S, and one each between L2S/Published and L1S/Pub. Combining two and three-way sharing, student groups shared more commonality with each other than they did with the Published writers. Total sharing was highest among L2S writers (36.94%), but for all groups, unshared items represented the highest proportion of top 20 usage, peaking among L1S writers, at 76.58%. In terms of the potential subject specificity of unshared items, this was highest among L1S writers, at just under 68% (nine items). By contrast, L2S writers showed only one potentially subject-specific item, representing around 5.7% of unshared usage. Published writers also showed notably less subject specificity than L1S writers, at only 2 items (13.35%). One explanation for this could be the purpose of writing in these different groups. For students, part of the aim is to demonstrate knowledge and to convince teaching staff that they are familiar with the subject. Including bundles (although they may not consciously think of them as such) that directly reference subject terminology or concepts may be one strategy to achieve this. For published writers, who are already established in their subject area, this may be of less importance. Lower subject specificity among L2S students may be a result of the additional difficulties of writing in a second language – other considerations take priority. Shared bundles in this discipline did not show any subject specificity.

4.4.2.2(ii) Education

Table 63a-20 most frequently used 4-word bundles for each writer group in Education

L2S	F	L1S	F	Published	F
students with special needs	49	for children with SEN	15	in the context of	37
the ministry of education	34	with special educational needs	12	the Ministry of Education	32
will be able to	23	to be able to	11	the ways in which	28
of inclusive education in	23	of children with SEN	10	of students with SEN	27
for children with SEN	23	it is important to	**10**	of children with disabilities	24
on the other hand	20	working with children with	8	of students with disabilities	23
children with special needs	20	with children with SEN	8	at the same time	23
I will be able	17	when it comes to	8	at the end of	22
to make sure that	13	the rest of the	8	students with special needs	21
it is important to	12	special educational needs and	8	of inclusive education in	21
to students with special	11	in order to achieve	8	the number of students	20
students with learning disabilities	11	i am going to	8	as well as the	20
of children with SEN	11	children with SEN to	8	it is important to	19
children with SEN in	11	the other children in	7	implementation of inclusive education	19
will help me to	10	in a mainstream school	7	on the basis of	18
when it comes to	10	children with SEN in	7	in the United States	18
i will be using	10	with the other children	6	for children with disabilities	18
children with SEN are	10	with the aim of	6	to the development of	17
as a result of	10	to meet the needs	6	their typically developing peers	17
to support children with	9	the needs of all	6	on the other hand	17

Table 63b-Top 20 4-word bundles item sharing between writer groups in Education

	L2S	L1S	Published
Total frequency / % of total top 20 usage	337 / 100	167 / 100	441 / 100
Shared between all groups frequency / % of total top 20 usage	12 / 3.56	10 / 5.99	19/ 4.30
Shared between L2S/L1S frequency / % of total top 20 usage	44 / 13.06	33 / 19.76	n/a
Shared between L2S/Published frequency / % of total top 20 usage	125 / 37.39	n/a	91 / 20.64
Shared between L1S/Published frequency / % of total top 20 usage	n/a	0/ 0	0 / 0
Total shared % of total top 20 usage	54.01	25.75	24.94
Unshared frequency / % of total top 20 usage	155 / 45.99	124 / 74.25	331 / 75.06

It is important to was the only top 20 bundle shared amongst all three groups. L2S/Published

writers shared four items, and L2S/L1S writers three. While L1S writers shared more overall

with their L1S counterparts than they did the Published group, L2S writers usage bore more

similarity to that of Published writers than to the L1S group, with L2S/Published shared items

being generally higher frequency. In this case, total sharing constituted almost 55% of L2S top 20 usage, but only around 25% for L1S and Published writers. Potential subject specificity was again highest among L1S writers (11 items/~67%) and was at a similar level for L2S and Published writers (six items/~46.5% and seven items/~44.7% respectively). Of the eight different items shared in total, five could be said to be potentially subject specific to this discipline.

4.4.2.2(iii) Fashion

Table 64a-20 most frequently used 4-word bundles for each writer group in Fashion

L2S	F	L1S	F	Published	F
as can be seen	15	it is important to	**28**	in the context of	49
can be seen from	14	can be seen in	22	on the other hand	33
can be seen in	13	as well as the	21	at the same time	29
when it comes to	12	this is due to	19	per cent of the	28
at the same time	11	the range will be	15	it is important to	28
for the reason that	10	it could be argued	15	is positively related to	27
turns out to be	8	is one of the	15	the purpose of this	26
the range will be	8	at the end of	15	in the case of	26
the performance of the	8	in line with the	14	the results of the	24
reasonable to think that	8	the end of the	13	one of the most	24
one of the most	8	the cost of the	12	as well as the	24
it is reasonable to	8	it is also important	12	on the basis of	23
is reasonable to think	8	is due to the	12	the degree to which	22
is one of the	8	in the fashion industry	12	are more likely to	22
it is important to	6	could be argued that	12	the relationship between brand	21
the aim of this	5	as well as a	12	the nature of the	21
it needs to be	5	as can be seen	12	is one of the	21
in the fashion industry	5	within the fashion industry	11	the extent to which	19
despite the fact that	5	when it comes to	11	on the one hand	18
can be divided into	5	the quality of the	11	purpose of this paper	17

Table 64b-Top 20 4-word bundles item sharing between writer groups in Fashion

	L2S	L1S	Published
Total frequency / % of total top 20 usage	170 / 100	294/ 100	502 / 100
Shared between all groups frequency / % of total top 20 usage	14 / 8.24	43 / 14.63	49 / 9.76
Shared between L2S/L1S frequency / % of total top 20 usage	45 / 26.47	57 / 19.39	n/a
Shared between L2S/Published frequency / % of total top 20 usage	19 / 11.18	n/a	53 / 10.56
Shared between L1S/Published frequency / % of total top 20 usage	n/a	21 / 7.14	24 / 4.78
Total shared % of total top 20 usage	45.88	41.16	25.10
Unshared frequency / % of total top 20 usage	92 / 54.12	173 / 58.84	376 / 74.90

Two bundles occurred in the top 20 for all three writer groups. L2S/Published writers shared four items, L2S/Published writers two, and L1S/Published writers only one item. Once again, if sharing overall is considered, student writers from both groups showed more commonality with each other than with the Published group. Published writers showed the highest proportion of unshared items, perhaps indicating a more varied repertoire of frequently used bundles. Potential subject specificity was generally lower than in Advertising or Education, but once again the highest level was found among L1S writers, at ~15% (two items), with Published writers showing the lowest level (one item or 5.59%). Of the nine different bundles shared in total, only one was subject specific. These relatively low levels would seem to concur with those from the analysis of writer groups across disciplines, in which subject specificity in Fashion was also generally rather low.

4.4.2.2(iv) Forensic Psychology

Table 65a-20 most frequently used 4-word bundles for each writer group in Forensic Psychology

L2S	F	L1S	F	Published	F
that Mr Taylor has	9	the criminal justice system	83	were more likely to	34
in the case of	7	vulnerable victims and witnesses	56	more likely to be	30
in an attempt to	7	it is important to	53	are more likely to	27
as well as the	7	present and highly relevant	42	the extent to which	26
it is important to	6	as a result of	42	the presence of a	24
in England and Wales	6	in England and Wales	41	as a result of	22
as a result of	6	are more likely to	41	the purpose of this	21
the Reid technique is	5	the way in which	36	it is important to	21
the criminal justice system	5	relationship with his mother	36	in the context of	21
reduce the risk of	5	and Criminal Evidence Act	36	in the control group	20
police and criminal evidence	5	is likely to be	34	per cent of the	18
on the other hand	5	within the criminal justice	31	in the current study	17
it was found that	5	it was found that	31	at the time of	17
it is necessary to	5	Mr Taylor has a	29	it is possible that	16
it is crucial to	5	has been found to	29	truth tellers and liars	15
it has been found	5	at the time of	29	there was no significant	15
it has also been	5	were more likely to	28	in line with the	15
in the Reid technique	5	have been found to	28	this paper is to	14
in comparison to the	5	it is likely that	27	of the current study	14
and Criminal Evidence Act	5	research has found that	26	be more likely to	14

Table 65b-Top 20 4-word bundles item sharing between writer groups in Forensic Psychology

	L2S	L1S	Published
Total frequency / % of total top 20 usage	113 / 100	758 / 100	401 / 100
Shared between all groups frequency / % of total top 20 usage	12 / 10.62	95 / 12.53	43 / 10.72
Shared between L2S/L1S frequency / % of total top 20 usage	21 / 18.58	191 / 25.20	n/a
Shared between L2S/Published frequency / % of total top 20 usage	0 / 0	n/a	0 / 0
Shared between L1S/Published frequency / % of total top 20 usage	n/a	98 / 12.93	78 / 19.45
Total shared % of total top 20 usage	29.20	50.66	30.18
Unshared frequency / % of total top 20 usage	80 / 70.80	374 / 49.34	280 / 69.82

It is important to and *as a result of* were the only top 20 bundles shared between all three groups. L2S/L1S writers shared four bundles, while L1S/Published writers shared three. L2S and Published writers both showed a higher proportion of unshared usage compared to L1S writers and thus total sharing was highest among this latter group. With two and three-way

sharing combined, L2S and L1S writers, as in preceding disciplines, shared more with each other than with Published writers. Potential subject specificity among unshared items was highest among the L1S group at 40.64%. L2S writers showed four items (30%), while Published writers exhibited the lower level of one item (5.36%). Two of the nine different bundles shared overall appeared to be subject specific.

4.4.2.2(v) Languages

Table 66a-20 most frequently used 4-word bundles for each writer group in Languages

L2S	F	L1S	F	Published	F
effective teaching and learning	41	effective teaching and learning	58	at the same time	46
it is important to	23	the rest of the	24	in the context of	43
I was able to	23	of the target language	24	on the other hand	39
English as an additional	23	use of the target	19	the extent to which	32
as an additional language	23	rest of the class	19	it is important to	32
the rest of the	21	in the target language	19	the use of the	28
of the target language	20	in the MFL classroom	19	in the case of	27
in the target language	20	I was able to	18	in relation to the	27
to be able to	19	use of target language	17	as well as the	27
I would like to	19	to effective teaching and	16	the ways in which	26
the use of the	18	English as an additional	15	the end of the	24
one of the most	16	as well as the	15	at the beginning of	21
at the same time	15	as an additional language	15	of the target language	20
use of the target	13	my block a placement	12	in the process of	20
use of target language	13	the use of target	11	in terms of the	19
when it comes to	11	the use of the	10	as a foreign language	19
to effective teaching and	11	it is important to	10	the results of the	18
the end of the	11	at the start of	10	in the target language	18
on the other hand	11	teaching and learning in	9	the language of the	17
as well as the	11	of effective teaching and	9	on the one hand	17

Table 66b-Top 20 4-word bundles item sharing between writer groups in Languages

	L2S	L1S	Published
Total frequency / % of total top 20 usage	362 / 100	249 / 100	520 / 100
Shared between all groups frequency / % of total top 20 usage	92 / 25.41	78 / 22.35	109 / 20.96
Shared between L2S/L1S frequency / % of total top 20 usage	157 / 43.37	166 / 47.56	n/a
Shared between L2S/Published frequency / % of total top 20 usage	37 / 10.21	n/a	109 / 20.96
Shared between L1S/Published frequency / % of total top 20 usage	n/a	0 / 0	0 / 0
Total shared % of total top 20 usage	79.01	69.91	55.00
Unshared frequency / % of total top 20 usage	76/ 20.99	105 / 30.09	286 / 45.00

Five top 20 bundles were shared between all three groups, while L2S/L1S writers shared seven, and L2S/Published writers a further three. Considering overall sharing, again the student groups exhibited a greater degree of mutual commonality compared to their sharing with the Published writers. Total sharing was highest among L2S writers and lowest in the Published group. Potential subject specificity was high for both student groups, with L1S showing seven items (90.48% of unshared usage) and L1S somewhat lower with three items (67.10%). Published were again the lowest in this regard (three items/12.59%). Seven of the 15 different shared bundles overall showed subject specificity.

4.4.2.2(vi) Physiotherapy

Table 67a-20 most frequently used 4-word bundles for each writer group in Physiotherapy

L2S	f	L1S	f	Published	f
-		it could be argued	7	of this study was	29
-		quality and compassionate care	6	this study was to	26
-		it is important to	**6**	in the intervention group	25
-		high quality and compassionate	6	times more likely to	18
-		the NHS leadership academy	4	in patients with chronic	18
-		to the lack of	3	this is the first	17
-		the national health service	3	the purpose of this	15
-		staff adopt leadership roles	3	see online supplementary material	15
-		national health service NHS	3	has been shown to	15
-		focused on individual leader	3	purpose of this study	14
-		due to the lack	3	the end of the	13
-		could be argued that	3	body mass index BMI	13
-		as part of a	3	is the first study	12
-		and quality of care	3	at the time of	12
-		all staff adopt leadership	3	at the end of	12
-		-		the results of this	11
-		-		it is possible that	11
-		-		in the present study	11
-		-		as a result of	11
-		-		with the exception of	10

Due to the small size of the Physiotherapy L2S sub-corpus, no bundles were identified, and only 15 were generated from the L1S sub-corpus. No sharing between the L1S and Published groups occurred. Subject specificity among the items was higher among the L1S group, with nine items (57.63%). Published writers showed only three items (18.18%) that exhibited potential subject specificity.

4.4.2.2(vii) Social Work

Table 68a-20 most frequently used 4-word bundles for each writer group in Social Work

L2S	F	L1S	F	Published	F
on the other hand	9	in the case study	33	child abuse and neglect	42
it is important to	9	it is important to	**26**	in the context of	41
in the case study	9	the service user to	23	the extent to which	31
for social workers to	7	for social workers to	18	it is important to	31
to the case study	5	the social worker and	17	in relation to the	29
to social work practice	5	of the life course	17	as a result of	26
the service user to	5	at the centre of	16	on the other hand	25
have an impact on	5	to the service user	15	in the United States	24
is important to consider	4	the social worker to	15	at the same time	24
that there is a	3	the service user and	15	as well as the	24
that the service user	3	are more likely to	15	of social work in	20
take into account the	3	a family group conference	15	are more likely to	20
of the social worker	3	to the case study	14	to be able to	18
of the limitations of	3	the solution focussed approach	13	the development of a	18
essay will discuss the	3	on the other hand	13	to the development of	17
as it is a	3	in social work practice	13	more likely to be	17
-		to social work practice	12	of social work practice	16
-		the service user is	12	in social work practice	16
-		the life course is	12	at the time of	16
-		from the case study	12	of child abuse and	15

Table 68b-Top 20 4-word bundles item sharing between writer groups in Social Work

	L2S	L1S	Published
Total frequency / % of total top 20 usage	79 / 100	326 / 100	470 / 100
Shared between all groups frequency / % of total top 20 usage	18 / 22.79	39 / 11.96	56 / 11.92
Shared between L2S/L1S frequency / % of total top 20 usage	31 / 39.24	100 / 30.68	n/a
Shared between L2S/Published frequency / % of total top 20 usage	0 / 0	n/a	0 / 0
Shared between L1S/Published frequency / % of total top 20 usage	28 / 8.59	0 / 0	36 / 7.66
Total shared % of total top 20 usage	62.02	51.23	19.57
Unshared frequency / % of total top 20 usage	30 / 37.98	159 / 48.77	378 / 80.43

It is important to and *on the other hand*, the two most commonly used bundles in the corpus as a whole, were shared between all three writer groups in this discipline. L2S/L1S and L1S/Published sharing amounted to five and two bundles respectively. Combining two and three-way sharing, for both student groups, sharing with each other was more widespread

than sharing with Published writers. Total sharing was highest in the L2S group, and in comparison, Published writers shared less than 20% of their top 20 usage. L1S writers, as in the other disciplines, showed the highest level of potential subject specificity, with nine items or 82.39% of unshared usage. By comparison the other groups were considerably lower – L2S showing two items (20%) and Published showing four items (24.6%). As for specificity among shared items, of the nine different bundles shared in this discipline, four could be said to be potentially subject specific.

4.4.2.2(viii) TESOL

Table 69a-20 most frequently used 4-word bundles for each writer group in TESOL

L2S	F	L1S	F	Published	F
on the other hand	78	will be able to	63	the extent to which	40
at the end of	52	at the end of	49	on the other hand	40
the end of the	41	to be able to	46	on the basis of	39
is one of the	34	the end of the	46	it is important to	38
at the same time	28	when it comes to	36	the end of the	37
in the field of	27	the results of the	33	in the United States	37
as a foreign language	25	as a result of	32	in the context of	36
in the target language	23	will be used to	31	in the present study	34
to be able to	22	it is important to	30	at the same time	31
this study aims to	21	in the form of	24	as a foreign language	30
the purpose of the	21	in the context of	24	English as a foreign	29
native speakers of English	21	through the use of	23	the use of the	28
it is important to	21	students will be able	22	in the form of	28
end of the course	21	is one of the	22	at the end of	28
will be asked to	20	in the field of	22	the ways in which	25
one of the most	20	of the English language	21	as a result of	24
level of English proficiency	20	the purpose of this	20	the start of the	23
at the beginning of	20	in the case of	20	as well as the	23
will be able to	19	in line with the	20	at the beginning of	22
when it comes to	19	as well as the	20	the beginning of the	21

Table 69b-Top 20 4-word bundles item sharing between writer groups in TESOL

	L2S	L1S	Published
Total frequency / % of total top 20 usage	553 / 100	604 / 100	613 / 100
Shared between all groups frequency / % of total top 20 usage	114 / 20.62	125 / 20.70	103 / 16.80
Shared between L2S/L1S frequency / % of total top 20 usage	121 / 21.88	189 / 31.29	n/a
Shared between L2S/Published frequency / % of total top 20 usage	151 / 27.31	n/a	123 / 20.07
Shared between L1S/Published frequency / % of total top 20 usage	100 / 16.56	0 / 0	111 / 18.12
Total shared % of total top 20 usage	69.80	68.54	54.98
Unshared frequency / % of total top 20 usage	167 / 30.20	190 / 31.46	276 / 45.02

Along with *it is important to*, the bundles *at the end of* and *the end of the* were shared between all three groups. Five bundles were shared between L2S/L1S, four between L2S/Pub, and four between L1S/Pub. As in most of the other disciplines, with two and three-way sharing combines, student writers bore more similarity to each other in terms of their top 20 lexical bundle use, than either of them did to the Published group. In total, the student groups shared almost 70% of their top 20 usage, while this figure was lower for Published writers. TESOL was the only discipline in which subject specificity in unshared items was highest for L2S writers (four items/50.9%). This dropped to 22.63% (two items) for L1S and 10.51% (one item) for Published writers.

4.4.2.2(ix) Summary of Within-Disciplines Results

To summarise the results within disciplines, Languages showed the most items shared between all writer groups (five). The bundle *it is important to* was shared among all groups in six of the eight disciplines. In terms of sharing overall, that between L2S and L1S writers was generally more extensive than sharing between student groups and Published writers. This indicates that, where the most commonly used bundles are concerned, Published writers' usage tends to differ from that of both L2 and L1 students, while the student writers themselves utilise a relatively more similar selection of bundles. The only disciplines in which

this was not the case were Physiotherapy (where no sharing occurred and thus no picture could be drawn) and Education, where more items and a greater proportion of usage were shared between L2S/Published writers. Total sharing was highest among the L2S group in most disciplines, with the exception of Physiotherapy, and Forensic Psychology, in which the L1S group had the highest level of sharing. Where sharing occurred, there were only two disciplines in which Published writers did not have the lowest total sharing; in Forensic Psychology sharing was lowest among L2S writers, and in Advertising among L1S writers. In the majority of the disciplines then, it appears that Published writers are making more frequent use of bundles not utilised by their student counterparts. Subject specificity in unshared items was highest among L1S writers in all disciplines, with the exception of TESOL.

4.4.2.3 Lexical Hedge use Within Disciplines

Tables 70a-77a show the 10 most frequent lexical hedges, with frequencies (F) and percentage of total hedge usage for the three writer groups within each of the disciplines in turn. Sharing is indicated using the same system outlined in previous within-disciplines analyses. Tables 70b-77b show the extent of item sharing between writer groups in the form of percentages of the total usage within the top 10 items.

4.4.2.3(i) Advertising

Table 70a-10 most frequently used lexical hedges for each writer group in Advertising

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
could	88	19.13	could	178	14.66	may	571	17.03
often	50	10.87	may	140	11.53	suggest-	333	9.93
might	37	8.04	suggest-	131	10.79	could	252	7.52
likely	33	7.17	often	90	7.41	indicate-	248	7.40
suggest-	27	5.87	indicate-	60	4.94	likely	246	7.34
argue-	24	5.22	argue-	58	4.78	might	175	5.22
may	19	4.13	likely	48	3.95	often	150	4.47
seem-	16	3.18	feel-	41	3.78	tend to-	106	3.16
possible -	14	3.04	seem-	39	3.21	seem-	104	3.10
tend- to	14	3.04	generally	31	2.55	in general	97	2.89
Total	322	69.69	Total	816	67.6	Total	2282	68.06

Table 70b-Top 10 lexical hedges item sharing between writer groups in Advertising

	L2S	L1S	Published
Total frequency / % of total top 10 usage	322/100	816 / 100	2282/ 100
Shared between all groups frequency / % of total top 10 usage	233/72.36	626/76.72	1656/72.57
Shared between L2S/L1S frequency / % of total top 10 usage	24/7.45	58/7.11	n/a
Shared between L2S/Published frequency / % of total top 10 usage	51/15.84	n/a	281/12.31
Shared between L1S/Published frequency / % of total top 10 usage	n/a	60/7.35	248/10.87
Total shared % of total top 10 usage	95.65	91.18	95.75
Unshared frequency / % of total top 10 usage	14/4.35	72/8.82	97/4.25

The top ten items made up a similar percentage of total usage in each writer group, with L2S slightly higher at 69.69%. Six items (accounting for over 72% of top ten usage) were shared between all three groups, notably the modals *could* and *may*, the adverb *often*, and the lexical verbs *suggest* and *seem*. With two and three-way sharing combined, both student groups bore more similarity to the Published writers than they did to each other. Total sharing was lowest in the L1S group, but even here the level was over 91%, and only two items were not shared

with at least one other group, suggesting that common hedge usage was fairly uniform across writer groups in this discipline.

4.4.2.3(ii) Education

Table 71a-10 most frequently used lexical hedges for each writer group in Education

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
may	120	16.97	may	112	22.67	may	345	12.49
might	92	13.01	often	62	12.55	suggest-	238	8.61
could	61	8.63	could	52	10.53	could	200	7.24
often	30	4.24	appear-	30	6.07	often	189	6.84
tend- to	28	3.96	argue-	30	6.07	indicate-	189	6.84
claim-	26	3.68	feel-	29	5.87	might	165	5.97
mainly	25	3.54	suggest-	27	5.47	argue-	131	4.74
argue-	24	3.39	likely	19	3.85	likely	86	3.11
indicate-	23	3.25	seem-	13	2.63	seem-	101	3.66
usually	21	2.97	usually	9	1.82	possible	75	2.71
Total	450	63.64	Total	383	77.53	Total	1719	62.21

Table 71b-Top 10 lexical hedges item sharing between writer groups in Education

	L2S	L1S	Published
Total frequency / % of total top 10 usage	450/100	383/100	1719/ 100
Shared between all groups frequency / % of total top 10 usage	235/52.22	256/66.84	865/50.32
Shared between L2S/L1S frequency / % of total top 10 usage	21/4.67	9/2.35	n/a
Shared between L2S/Published frequency / % of total top 10 usage	115/25.57	n/a	354/20.59
Shared between L1S/Published frequency / % of total top 10 usage	n/a	59/15.40	425/24.72
Total shared % of total top 10 usage	82.44	84.60	95.64
Unshared frequency / % of total top 10 usage	79/17.56	59/15.40	75/4.36

While the top ten hedges accounted for a similar proportion of total usage in the L2S and Published groups, this figure was higher among L1S writers (77.53%), perhaps indicating a heavier reliance on the most frequent forms in this group. Only four hedges were common to all three groups, but these again included the modals *could* and *may*, and the adverb *often*. Overall, sharing between the L2S and L1S writers was notably lower than sharing between either L2S and Published or L1S and Published groups, indicating that, where frequent hedge

use is concerned, both L2S and L1S have more in common with Published writers than they do with each other. Lowest overall sharing was found in the L2S group, with three unshared items – *claim*, *tend to*, and *mainly*. Published writers showed the highest total sharing.

4.4.2.3(iii) Fashion

Table 72a-10 most frequently used lexical hedges for each writer group in Fashion

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
could	72	18.56	could	186	16.67	may	559	18.92
may	29	7.47	may	164	14.77	suggest-	286	9.68
seem- -	26	6.70	suggest-	88	7.93	could	185	6.26
estimate-	23	5.93	indicate-	62	5.59	often	176	5.96
suggest-	22	5.67	likely	61	5.50	likely	175	5.92
might	19	4.90	estimate-	56	5.05	indicate-	169	5.72
likely	16	4.12	often	43	3.87	might	124	4.20
claim-	14	3.61	seem-	34	3.06	estimate-	123	4.16
indicate-	13	3.35	around	31	2.79	argue-	94	3.18
around	10	2.58	typically -	28	2.52	tend- to	79	2.67
quite	10	2.58						
roughly	10	2.58						
usually	10	2.58						
Total	274	70.63	Total	753	67.75	Total	1970	66.67

Table 72b-Top ten lexical hedges item sharing between writer groups in Fashion

	L2S	L1S	Published
Total frequency / % of total top 10 usage	274/100	753/100	1970/ 100
Shared between all groups frequency / % of total top 10 usage	175/63.87	617/81.94	1497/75.99
Shared between L2S/L1S frequency / % of total top 10 usage	36/13.14	65/8.63	n/a
Shared between L2S/Published frequency / % of total top 10 usage	19/6.93	n/a	124/6.29
Shared between L1S/Published frequency / % of total top 10 usage	n/a	43/5.71	176/8.93
Total shared % of total top 10 usage	83.94	96.28	91.22
Unshared frequency / % of total top 10 usage	44/16.06	28/3.72	173/8.78

The proportion of total hedge usage constituted by the top ten was similar in the three groups (although this appears higher in the L2S group, the 'top 10' by frequency in this case was actually a top 13, and this must be factored in when considering the percentages). Six items were common to all three groups, including once again the modals *could* and *may*, and the

213

lexical verb *suggest*. *Claim* once again features among the unshared items for L2S writers. With two and three-way sharing considered together, the student groups showed greater commonality with each other than with the Published writers. Total sharing was highest in the L1S group.

4.4.2.3(iv) Forensic Psychology

Table 73a-10 most frequently used lexical hedges for each writer group in Forensic Psychology

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
may	69	16.67	may	935	23.76	may	604	21.11
could	44	10.63	suggest-	577	14.66	suggest-	275	9.61
suggest-	39	9.42	could	460	11.69	likely	259	9.05
likely	25	6.04	likely	407	10.34	indicate-	189	6.61
indicate-	24	5.80	often	221	5.62	could	183	6.40
tend- to	20	4.83	indicate-	159	4.04	might	139	4.86
might	15	3.62	seem-	135	3.34	often	130	4.54
often	14	3.38	appear-	122	3.10	appear-	76	2.66
argue-	13	3.14	feel-	121	3.07	possible	67	2.34
usually	12	2.90	argue-	103	2.62	generally	57	1.99
claim-	12	2.90						
Total	287	69.33	Total	3240	82.24	Total	1979	69.17

Table 73b-Top 10 lexical hedges item sharing between writer groups in Forensic Psychology

	L2S	L1S	Published
Total frequency / % of total top 10 usage	287/100	3240/100	1979/ 100
Shared between all groups frequency / % of total top 10 usage	215/74.91	2759/85.15	1640/82.87
Shared between L2S/L1S frequency / % of total top 10 usage	12/4.53	103/3.18	n/a
Shared between L2S/Published frequency / % of total top 10 usage	15/5.23	n/a	139/7.02
Shared between L1S/Published frequency / % of total top 10 usage	n/a	122/3.77	76/3.84
Total shared % of total top 10 usage	84.67	92.10	93.73
Unshared frequency / % of total top 10 usage	44/15.33	256/7.90	124/6.27

The proportion of total hedge usage represented by the top ten hedges was notably higher in the L1S group (82.24%). *May*, *could*, *suggest* and *often* occurred once more among the six items common to all three groups. Two way sharing was limited to one item in each category,

but considering both two and three-way sharing, the student groups both shared more commonality with Published writers than with each other. Total sharing was highest in the Published group. L2S unshared hedges again included *claim* and *tend to*, while not for the first time, *feel* occurred as an unshared item in the L1S group.

4.4.2.3(v) Languages

Table 74a-10 most frequently used lexical hedges for each writer group in Languages

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
may	106	12.09	may	122	14.56	may	529	16.92
might	83	9.46	often	82	9.79	suggest-	294	9.40
seem-	83	9.46	could	81	9.67	could	171	5.47
often	76	8.67	feel-	62	7.40	might	168	5.37
could	74	8.44	suggest-	52	6.21	indicate-	164	5.24
usually	45	5.13	seem-	46	5.49	often	157	5.02
suggest-	33	3.76	perhaps	32	3.82	seem-	157	5.02
sometimes	32	3.65	likely	29	3.46	argue-	141	4.19
in my opinion	24	2.74	argue-	29	3.46	appear-	119	3.81
argue-	24	2.74	appear-	26	3.10	likely	94	3.01
Total	580	66.14	Total	561	66.96	Total	1994	63.45

Table 74b-Top 10 lexical hedges item sharing between writer groups in Languages

	L2S	L1S	Published
Total frequency / % of total top 10 usage	580/100	561/100	1994/ 100
Shared between all groups frequency / % of total top 10 usage	396/68.28	412/73.44	1449/72.67
Shared between L2S/L1S frequency / % of total top 10 usage	0/0	0/0	n/a
Shared between L2S/Published frequency / % of total top 10 usage	83/14.31	n/a	168/8.43
Shared between L1S/Published frequency / % of total top 10 usage	n/a	55/9.80	213/10.68
Total shared % of total top 10 usage	82.59	83.24	91.78
Unshared frequency / % of total top 10 usage	101/17.41	94/16.76	164/8.22

While the proportions of total hedge use were similar, top 10 usage accounted for a slightly smaller proportion in the Published writer group (63.45%). Once again, six items were common to all groups, including *may, could, suggest* and *often*. Overall, both L2S and L1S writers shared more with the Published group than with each other. L1S unshared items included *feel*, as was

the case in a number of the other disciplines. Total sharing was highest among the Published group.

4.4.2.3(vi) Physiotherapy

Table 75a-10 most frequently used lexical hedges for each writer group in Physiotherapy

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
may	14	19.18	suggest-	51	20.40	may	305	20.76
suggest-	9	12.33	may	46	18.40	indicate-	149	10.14
might	7	9.59	could	44	17.60	could	148	10.07
argue- -	7	9.59	often	18	7.20	suggest-	141	9.60
claim-	6	8.22	argue-	14	5.60	likely	98	6.67
often	5	6.85	appear-	10	4.00	estimate-	80	5.45
postulate-	5	6.85	tend- to	6	2.40	might	72	4.90
could	3	4.11	assume-	6	2.40	often	49	3.34
seem-	3	4.11	generally	5	2.00	approximately	45	3.06
feel-	2	2.74	likely	5	2.00	possible	39	2.65
sometimes	2	2.74	indicate-	5	2.00			
Total	63	86.31	Total	210	84.00	Total	1126	76.64

Table 75b-Top 10 lexical hedges item sharing between writer groups in Physiotherapy

	L2S	L1S	Published
Total frequency / % of total top 10 usage	63/100	210/100	1126/ 100
Shared between all groups frequency / % of total top 10 usage	31/49.21	159/75.71	641/57.11
Shared between L2S/L1S frequency / % of total top 10 usage	7/11.11	14/6.68	n/a
Shared between L2S/Published frequency / % of total top 10 usage	7/11.11	n/a	72/6.39
Shared between L1S/Published frequency / % of total top 10 usage	n/a	10/4.76	247/21.94
Total shared % of total top 10 usage	71.43	87.14	85.44
Unshared frequency / % of total top 10 usage	18/28.57	27/12.86	164/14.56

The Published group showed the lowest proportion of total hedge usage accounted for by the ten most common hedges. Item sharing between all three groups was somewhat lower in this discipline, with only four items – *could, may, suggest* and *often*. Considering overall sharing between both two and three groups, L2S writers shared an equal proportion of their top 10 usage with the L1S and Published groups, while the L1S writers themselves showed usage more

similar to that of their student counterparts. *Claim* appeared again as an unshared item among

L2S writers, who also showed the lowest total sharing, at only 71.43%.

4.4.2.3(vii) Social Work

Table 76a-10 most frequently used lexical hedges for each writer group in Social Work

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
could	69	29.36	may	249	27.01	may	528	17.50
might	23	9.79	could	164	17.79	suggest-	257	8.52
suggest-	18	7.66	suggest-	149	16.16	might	232	7.69
may	15	6.38	argue-	67	7.27	likely	192	6.36
argue-	15	6.38	likely	50	5.42	indicate-	187	6.20
appear-	14	5.96	often	33	3.58	often	182	6.03
claim-	9	3.83	feel-	23	2.49	could	140	4.64
possibly	7	2.98	sometimes	19	2.06	argue-	121	4.01
possible	6	2.55	appear-	19	2.06	tend- to	84	2.78
often	6	2.55	might	18	1.95	seem-	75	2.49
indicate-	6	2.55						
Total	188	79.99	Total	791	85.79	Total	1998	66.22

Table 76b-Top 10 lexical hedges item sharing between writer groups in Social Work

	L2S	L1S	Published
Total frequency / % of total top 10 usage	188/100	791/100	1998/ 100
Shared between all groups frequency / % of total top 10 usage	146/77.67	680/85.97	1460/57.11
Shared between L2S/L1S frequency / % of total top 10 usage	14/7.45	19/2.40	n/a
Shared between L2S/Published frequency / % of total top 10 usage	6/3.19	n/a	187/9.36
Shared between L1S/Published frequency / % of total top 10 usage	n/a	50/6.32	192/9.61
Total shared % of total top 10 usage	88.30	94.69	92.04
Unshared frequency / % of total top 10 usage	22/11.70	42/5.31	159/7.96

Top ten hedges as a proportion of total usage were notably lower among Published writers.

The four items that have commonly occurred as shared between all groups in a number of

other disciplines (*could, may, suggest, often*) did so again here. Additionally, *claim* reappeared

as an unshared item for L2S writers, as did *feel* in the L1S group. With two and three-way

sharing considered together, L2S writers showed greatest commonality with the other student

group, while L1S writers' usage most closely mirrored that of the Published group. Total sharing was highest among L1S writers, at 94.69%.

4.4.2.3(viii) TESOL

Table 77a-10 most frequently used lexical hedges for each writer group in TESOL

L2S			L1S			Published		
Hedge	F	%	Hedge	F	%	Hedge	F	%
might	390	15.42	may	585	16.28	may	638	15.18
may	279	11.03	suggest-	524	14.58	suggest-	367	8.73
could	268	10.60	could	399	11.10	indicate-	265	6.31
suggest-	231	9.13	often	214	5.96	often	245	5.83
seem-	130	5.14	argue-	123	3.42	might	245	5.83
argue-	116	4.59	seem-	119	3.41	likely	192	4.57
indicate-	90	3.56	likely	101	2.81	seem-	168	4.00
often	76	3.01	appear-	101	2.81	could	167	3.97
claim-	76	3.01	indicate	89	2.48	argue-	148	3.52
tend- to	68	2.69	possible	87	2.42	appear-	109	2.59
			feel-	87	2.42			
Total	1724	68.18	Total	2429	67.69	Total	2544	60.53

Table 77b-Top 10 lexical hedges item sharing between writer groups in TESOL

	L2S	L1S	Published
Total frequency / % of total top 10 usage	1724/100	2429/100	2554/ 100
Shared between all groups frequency / % of total top 10 usage	1190/69.03	2053/84.52	1998/78.54
Shared between L2S/L1S frequency / % of total top 10 usage	0/0	0/0	n/a
Shared between L2S/Published frequency / % of total top 10 usage	390/22.62	n/a	245/9.63
Shared between L1S/Published frequency / % of total top 10 usage	n/a	202/8.31	301/11.83
Total shared % of total top 10 usage	91.65	92.84	100
Unshared frequency / % of total top 10 usage	144/8.35	174/7.16	0/0

Published writers showed the lowest proportion of total hedges accounted for by the top ten. In this discipline, seven items were common to all groups, and again these included *could, may*, *suggest* and *often*. *Claim* and *feel* appear again as unshared items in the L2S and L1S groups respectively. In terms of sharing overall, both student groups showed more commonality with

the Published group than with each other. Total sharing was very high in this discipline, being over 91% in all groups, and peaking at 100% among Published writers.

4.4.2.3(ix) Summary of Within-Disciplines Results.

In summary, the proportion of total hedge usage accounted for by the top 10 was lowest among Published writers in seven of the eight disciplines (although the difference was minimal in some cases), suggesting that this writer group relies less heavily on these most frequent items. Total sharing within the top ten was lowest among L2S writers in seven of the eight disciplines (the exception being Advertising). This may be an indication that the L2S group employ with greater frequency, hedges that may be relatively uncommon among the other two groups. The hedges *may, could* and *suggest*, were common to all three groups in every discipline, with *often* common to all in six disciplines, and *argue* common to all in four. The modal *might* was common to L2S/Published writers in six of the eight disciplines, but only appeared in the L1S top ten in one discipline (Social Work), and that at position 10. In terms of unshared items, *claim* made the top 10 lists for L2S writers in six disciplines, as did *feel* in the case of L1S writers. Published writers showed less consistency in unshared items, the most common being *possible*, although this only appeared in three disciplines. Considering sharing overall, in five of the eight disciplines (Advertising, Education, Forensic Psychology, Languages, and TESOL), both student groups showed usage more similar to that of the Published group than to each other. In the Fashion and Social Work disciplines, student groups resembled each other in usage to a greater extent than either resembled the Published writers, indicating that in these cases, the most frequently used hedges among the Published group show variance from those most commonly employed by students.

4.4.2.4 Citation Use within Disciplines

The following sections detail citation use within each of the eight disciplines. Firstly, there is

an overview of each discipline in terms of integral/non-integral citation use, and the use of the

various integral forms, and secondly an examination of reporting verb use and sharing

between the different writer groups in each discipline.

4.4.2.4(i) Overall Citation Use Within Disciplines

Tables 78-85 show rates of overall citation per 1000 words, as well as use of integral/non-

integral citations and the various integral forms, expressed as percentages of total integral

citation use for the three writer groups in each of the disciplines.

4.4.2.4(i)a Advertising

Table 78-Overall citation use in Advertising

	L2S	L1S	Published
Citations/1000 words	6.20	6.70	4.70
Non-integral citations %	67.53	63.91	51.23
Integral citations %	32.47	36.09	48.77
Non-Reporting integral citations %	18.00	14.41	22.51
According to %	29.00	4.41	6.20
Discourse act verbs %	33.00	37.94	33.77
Research act verbs %	17.00	36.18	32.61
Cognition act verbs %	3.00	7.06	3.03
Ambiguous %	0.00	0.00	1.86

Published writers showed the lowest overall citation rate, and the closest balance between

integral/non-integral use. Conversely, among L2S writers, non-integral citations were over

twice as common as integral citations. Non-reporting structures were slightly higher among

Published writers, while L2S writers made proportionally considerably more use of *according*

to. While discourse and cognition act verbs were used in broadly similar proportions by all

three groups, L2S writers exhibited proportionally lower use of research act verbs.

220

4.4.2.4(i)b Education

Table 79-Overall citation use in Education

	L2S	L1S	Published
Citations/1000 words	10.69	10.63	4.95
Non-integral citations %	78.48	74.72	55.32
Integral citations %	21.52	25.28	44.68
Non-Reporting integral citations %	12.74	8.96	25.00
According to %	20.59	0.75	3.74
Discourse act verbs %	46.57	57.46	41.33
Research act verbs %	12.75	13.43	24.32
Cognition act verbs %	7.35	19.40	5.27
Ambiguous %	0.00	0.00	0.34

The overall citation rate was lowest among Published writers, and although non-integral citations dominated in every group, this was most pronounced in the L2S cohort. Non-reporting structures were highest by proportion in the Published group, while L2S writers made notably more use of *according to.* All three groups relied most heavily on discourse act verbs, particularly in the case of L1S writers, although cognition act verb use was also proportionally higher in this group.

4.4.2.4(i)c Fashion

Table 80-Overall citation use in Fashion

	L2S	L1S	Published
Citations/1000 words	8.13	8.59	8.22
Non-integral citations %	86.37	88.65	76.59
Integral citations %	13.63	11.35	23.41
Non-Reporting integral citations %	30.50	8.84	27.00
According to %	35.59	20.40	6.32
Discourse act verbs %	20.34	51.02	35.90
Research act verbs %	10.17	15.65	27.86
Cognition act verbs %	3.39	3.40	2.39
Ambiguous %	0.00	0.68	0.51

Overall rates of citation were fairly consistent in this discipline, and although lowest among Published writers, the proportion of non-integral citations was particularly high across all groups. Non-reporting structures were relatively high by proportion in both the L2S and Published groups, but once again, *according to* represented a higher proportion of usage among L2S writers than among the other groups. All three writer groups favoured discourse act verbs, with L1S using them at a particularly high proportion.

4.4.2.4(i)d Forensic Psychology

Table 81-Overall citation use in Forensic Psychology

	L2S	L1S	Published
Citations/1000 words	12.60	11.67	7.44
Non-integral citations %	65.85	67.32	76.92
Integral citations %	34.15	32.68	23.08
Non-Reporting integral citations %	8.98	13.66	16.27
According to %	7.78	1.61	0.53
Discourse act verbs %	32.93	35.86	37.60
Research act verbs %	41.92	43.74	40.53
Cognition act verbs %	8.38	5.12	4.80
Ambiguous %	0.00	0.00	0.26

Published writers showed the lowest overall citation rate and the highest proportion of non-integral citations, although in all three groups these were more common than integral citations. Published writers also used proportionally more non-reporting forms than the student groups. Use of *according to* was once again highest among L2S writers. Research act verbs dominated in all three groups and at similar proportions.

4.4.2.4(i)e Languages

Table 82-Overall citation use in Languages

	L2S	L1S	Published
Citations/1000 words	4.19	4.05	4.31
Non-integral citations %	58.28	53.45	42.36
Integral citations %	41.72	46.55	57.64
Non-Reporting integral citations %	11.17	6.45	24.82
According to %	11.73	1.29	6.03
Discourse act verbs %	54.75	64.52	39.55
Research act verbs %	15.64	17.42	25.25
Cognition act verbs %	6.70	9.03	3.37
Ambiguous %	0.00	1.29	0.98

Published writers showed the highest rate of citation in this discipline, perhaps as a result of the fact that while the research articles in this discipline were obviously academic in nature, the students were participating in a course that was arguably more focused on vocational goals. Assignments therefore included extensive self-reflection from teaching placements, and this may have resulted in the use of fewer citations in general. Integral citations dominated usage among Published writers, while both student groups favoured non-integral forms. Non-reporting structures were highest by proportion in the Published group, while L2S writers made most use of *according to*. Discourse act verbs were predominant in every group, although most noticeably among the L1S writers.

4.4.2.4(i)f Physiotherapy

Table 83-Overall citation use in Physiotherapy

	L2S	L1S	Published
Citations/1000 words	8.65	9.51	10.78
Non-integral citations %	52.63	34.11	95.01
Integral citations %	47.37	65.89	4.99
Non-Reporting integral citations %	3.70	4.25	25.56
According to %	7.40	0.00	2.22
Discourse act verbs %	51.85	55.32	27.78
Research act verbs %	37.04	25.53	42.22
Cognition act verbs %	0.00	14.89	2.22
Ambiguous %	0.00	0.00	0

The overall citation rate was highest among Published writers, which may be a result of the numerical superscript system employed in the research articles, which also led to the unusually high proportion of non-integral citations in this group. Integral citations actually dominated among L1S writers. Non-reporting structures were most common among Published writers, while L2S writers made the most use of *according to.* Discourse act verbs were predominant in both student groups, while among Published writers, research act verbs made up the largest proportion of integral citation use. Cognition act verb use was notably higher in the L1S group.

4.4.2.4(i)g Social Work

Table 84-Overall citation use in Social Work

	L2S	L1S	Published
Citations/1000 words	18.60	16.95	8.10
Non-integral citations %	77.97	66.00	75.91
Integral citations %	22.03	34.00	24.09
Non-Reporting integral citations %	8.97	11.90	19.50
According to %	3.85	4.98	3.63
Discourse act verbs %	65.38	63.85	43.59
Research act verbs %	21.79	14.07	28.30
Cognition act verbs %	0.00	4.76	3.82
Ambiguous %	0.00	0.43	1.14

The lowest rate of overall citation was found in the Published group. Non-integral citations were used proportionally more in all three groups. Published writers made the most use by proportion of non-reporting forms, while the use of *according to* was similar across groups, being only slightly higher among L1S writers. All groups favoured discourse act verbs, although the difference in usage levels between these and research act verbs was less pronounced in the Published group.

4.4.2.4(i)h TESOL

Table 85-Overall citation use in TESOL

	L2S	L1S	Published
Citations/1000 words	12.43	8.38	6.94
Non-integral citations %	60.12	42.94	63.32
Integral citations %	39.88	57.06	36.68
Non-Reporting integral citations %	20.13	18.12	26.01
According to %	13.07	4.64	2.14
Discourse act verbs %	46.66	48.86	36.51
Research act verbs %	15.22	21.89	31.16
Cognition act verbs %	4.77	6.18	3.53
Ambiguous %	0.15	0.31	0.64

The overall citation rate was lowest among Published writers. L1S writers were the only group in which integral citations were more common than non-integral forms. The use of non-reporting structures was broadly balanced, although slightly higher among Published writers, while *according to* was used proportionally more by the L2S group. All groups utilised discourse act verbs more often than the other two verb types, although among Published writers, the difference between discourse act use and research act use was less pronounced.

4.4.2.4(ii) Reporting Verb Use and Sharing Within Disciplines

Tables 86a-93a show the 10 most frequent reporting verbs, with frequencies (F) and percentage of total integral citation use for the three writer groups within each of the

disciplines in turn. Sharing is indicated using the same system outlined in previous within-disciplines analyses. Tables 86b-93b show the extent of item sharing between writer groups in the form of percentages of the total usage within the top 10 items.

4.4.2.4(ii)a Advertising

Table 86a-10 most frequently used reporting verbs for each writer group in Advertising

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	29	29.00	find	45	13.24	suggest	51	7.36
state	7	7.00	state	31	9.12	find	48	6.93
suggest	4	4.00	suggest	17	5.00	according to	43	6.20
argue	3	3.00	according to	15	4.41	define	26	3.75
show	3	3.00	argue	11	3.24	propose	25	3.61
define	3	3.00	conduct -	10	2.94	develop	24	3.46
explain	3	3.00	explain	10	2.94	identify	23	3.32
find out	3	3.00	support	8	2.35	state	20	2.89
point out	2	2.00	describe	7	2.06	note	18	2.60
imply	2	2.00	conclude	5	1.47	show	15	2.16
believe	2	2.00	say	5	1.47	demonstrate	14	2.02
			claim	5	1.47			
Total %		61.00	Total %		49.71	Total %		44.30

Table 86b-Top 10 reporting verbs item sharing between writer groups in Advertising

	L2S	L1S	Published
Total frequency / % of total top 10 usage	61/100	169/100	307/100
Shared between all groups frequency / % of total top 10 usage	40/65.57	63/37.28	114/37.13
Shared between L2S/L1S frequency / % of total top 10 usage	6/9.84	21/12.43	n/a
Shared between L2S/Published frequency / % of total top 10 usage	6/9.84	n/a	41/13.36
Shared between L1S/Published frequency / % of total top 10 usage	n/a	45/26.63	48/15.64
Total shared % of total top 10 usage	85.25	76.33	66.12
Unshared frequency / % of total top 10 usage	9/14.75	40/23.67	104/33.88

The proportion of total integral citation usage represented by the ten most frequently used items was highest in the L2S group (61.00%), with the lowest proportion seen among Published writers (44.30%). This suggests that L2S writers in this discipline may be relying more heavily than their L1S counterparts and particularly the Published writers, on the most common

reporting verbs. Three items - *according to*, *state*, and *suggest* were shared between all three groups, making up a substantial 65.57% of L2S top 10 usage, and around 37% in the other groups. Both verbs here were discourse act verbs. Overall, if two and three-way sharing are combined, L2S writers shared the same proportion of top 10 usage with L1S writers as they did with Published writers, while the L1S group shared more similarity with the Published group. Total sharing was highest among L2S, and lowest among Published writers, again suggesting a more frequent use of varied reporting verbs in this latter group.

4.4.2.4(ii)b Education

Table 87a-10 most frequently used reporting verbs for each writer group in Education

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	42	20.59	argue	23	17.16	argue	40	6.8
state	17	8.33	state	16	11.94	describe	28	4.76
claim	12	5.88	believe	14	10.45	suggest	27	4.59
suggest	7	3.43	suggest	9	6.72	according to	22	3.74
describe	6	2.94	agree	4	2.99	find	22	3.74
argue	6	2.94	define	4	2.99	show	13	2.21
support	5	2.45	find	4	2.99	report	13	2.21
mention	5	2.45	claim	3	2.24	note -	12	2.04
define	5	2.45	discuss	3	2.24	identify	12	2.04
refer to	4	1.96	report	3	2.24	propose	11	1.87
explain	4	1.96	say	3	2.24	conduct	9	1.53
			uncover	3	2.24			
			view	3	2.24			
Total %		55.38	Total %		68.68	Total %		35.53

Table 87b-Top 10 reporting verbs item sharing between writer groups in Education

	L2S	L1S	Published
Total frequency / % of total top 10 usage	113/100	92/100	209/100
Shared between all groups frequency / % of total top 10 usage	13/11.50	32/34.78	67/32.06
Shared between L2S/L1S frequency / % of total top 10 usage	34/30.09	23/25.00	n/a
Shared between L2S/Published frequency / % of total top 10 usage	48/42.48	n/a	50/23.92
Shared between L1S/Published frequency / % of total top 10 usage	n/a	7/7.61	35/16.75
Total shared % of total top 10 usage	84.07	67.39	72.73
Unshared frequency / % of total top 10 usage	18/15.93	30/32.61	57/27.27

Published writers showed the lowest proportion of total integral citation use made up by the ten most frequently used reporting verbs, at 35.53%. In this discipline, L1S writers showed the greatest reliance on these most common forms (68.88%), although this may in part be due to multiple identical frequencies. Only two items – *suggest* and *argue* (both discourse verbs) were shared between all groups, and these represented over 32% of top ten usage in the L1S and Published groups, but only 11.50% in the L2S group. Combining two and three-way sharing, L2S and Published writers bore more similarity to each other than either of them did to L1S writers. Total sharing was highest in L2S and lowest in L1S.

4.4.2.4(ii)c Fashion

Table 88a-10 most frequently used reporting verbs for each writer group in Fashion

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	21	35.59	state	31	21.09	according to	37	6.32
state	3	5.08	according to	30	20.41	suggest	36	6.15
suggest	2	3.39	report	12	8.16	find -	32	5.47
report	2	3.39	suggest	11	7.48	argue	24	4.1
argue	2	3.39	indicate	4	2.72	show	19	3.25
say	1	1.69	discuss	4	2.72	report	18	3.08
predict	1	1.69	define	3	2.04	develop	15	2.56
mention	1	1.69	describe	3	2.04	point out	12	2.05
list	1	1.69	find	3	2.04	propose	12	2.05
introduce	1	1.69	identify	3	2.04	use	12	2.05
conduct	1	1.69	note	2	1.36	describe	12	2.05
campaign	1	1.69	outline	2	1.36			
advise	1	1.69	prove	2	1.36			
sustain	1	1.69	provide	2	1.36			
succeed	1	1.69						
take into account	1	1.69						
Total %		69.43	Total %		76.18	Total %		39.13

228

Table 88b-Top 10 reporting verbs item sharing between writer groups in Fashion

	L2S	L1S	Published
Total frequency / % of total top 10 usage	41/100	112/100	229/100
Shared between all groups frequency / % of total top 10 usage	25/60.98	53/47.322	91/39.74
Shared between L2S/L1S frequency / % of total top 10 usage	3/7.32	31/27.68	n/a
Shared between L2S/Published frequency / % of total top 10 usage	2/4.88	n/a	24/10.48
Shared between L1S/Published frequency / % of total top 10 usage	n/a	6/5.36	44/19.21
Total shared % of total top 10 usage	73.17	80.36	69.43
Unshared frequency / % of total top 10 usage	11/26.83	22/19.64	70/30.57

Top ten usage made up over 76% of total integral citation use among L1S students, compared to 69.43% for L2S and a mere 39.13% for Published writers. Three items – *according to, suggest*, and *report* (both verbs discourse related) were shared between all three groups, making up the largest proportion of sharing in each case. Both student groups bore more similarity to each other in terms of item sharing than either of them did to the Published group if two and three-way sharing are combined. The Published group themselves showed more commonality with L1S than with L2S writers. Total sharing was highest in the L1S group, and as in the preceding disciplines, lowest among Published writers.

4.4.2.4(ii)d Forensic Psychology

Table 89a-10 most frequently used reporting verbs for each writer groups in Forensic Psychology

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
find	40	23.95	find	230	21.82	find	40	10.67
according to	13	7.78	state	98	9.30	suggest	34	9.07
suggest	9	5.39	suggest	67	6.36	report	28	7.47
support	8	4.79	conduct	43	4.08	identify	19	5.07
discover	7	4.19	argue	26	2.47	describe	16	4.37
conduct	7	4.19	note	25	2.37	conduct	13	3.47
argue	5	2.99	describe	24	2.28	show	10	2.67
emphasise	5	2.99	conclude	19	1.80	examine	6	1.60
report -	5	2.99	identify	19	1.80	develop	6	1.60
show	5	2.99	support	19	1.80	conclude	6	1.60
Total %		62.25	Total %		54.08	Total %		47.59

Table 89b-Top 10 reporting verbs item sharing between writer groups in Forensic Psychology

	L2S	L1S	Published
Total frequency / % of total top 10 usage	104/100	570/100	178/100
Shared between all groups frequency / % of total top 10 usage	56/53.85	340/59.65	87/48.88
Shared between L2S/L1S frequency / % of total top 10 usage	13/12.50	45/7.90	n/a
Shared between L2S/Published frequency / % of total top 10 usage	10/9.62	n/a	38/21.35
Shared between L1S/Published frequency / % of total top 10 usage	n/a	43/7.54	35/19.66
Total shared % of total top 10 usage	75.96	75.09	89.89
Unshared frequency / % of total top 10 usage	25/24.04	142/24.91	18/10.11

Again the proportion of total integral citation use represented by the ten most frequently occurring reporting verbs was highest in the L2S group (62,25%), and lowest in the Published group (47.59%). *Find* was the most common item in all three groups. Three items were shared between all groups (*find, suggest, conduct*), again representing the largest proportion of sharing for each group. Unusually, two research acts verbs were among the three most commonly shared items, which may be related to the fact that this discipline is one of only two in which research act verbs were more common than discourse act verbs. Combining two and three-way sharing, both student writer groups resembled each other more than they did the Published group. Total sharing was highest in the Published group, with the student groups showing similar levels to each other.

4.4.2.4(ii)e Languages

Table 90a-10 most frequently used reporting verbs for each writer group in Languages

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	21	11.73	state	20	12.90	find	51	7.15
argue	14	7.82	argue	14	9.03	according to	43	6.04
state	12	6.70	suggest	11	7.10	argue	37	5.19
point out	10	5.59	describe	9	5.81	note	25	3.51
mention	8	4.47	highlight	6	3.87	suggest	20	2.81
suggest	6	3.35	support	5	3.23	report	20	2.81
explain	6	3.35	define	4	2.58	describe	19	2.66
show -	5	2.79	discuss	4	2.58	show	18	2.52
affirm	4	2.23	say	4	2.58	observe	13	1.82
establish	3	1.68	find	3	1.94	propose	13	1.82
carry out	3	1.68	believe	3	1.94			
write	3	1.68	advocate	3	1.94			
Total %		53.07	Total %		55.50	Total %		36.33

Table 90b-Top 10 reporting verbs item sharing between writer groups in Languages

	L2S	L1S	Published
Total frequency / % of total top 10 usage	95/100	86/100	259/100
Shared between all groups frequency / % of total top 10 usage	20/21.05	25/29.07	57/22.01
Shared between L2S/L1S frequency / % of total top 10 usage	12/12.63	20/23.26	n/a
Shared between L2S/Published frequency / % of total top 10 usage	26/27.37	n/a	61/23.55
Shared between L1S/Published frequency / % of total top 10 usage	n/a	12/13.95	70/27.03
Total shared % of total top 10 usage	61.05	66.28	72.59
Unshared frequency / % of total top 10 usage	37/38.95	29/33.72	71/27.41

Top 10 items constituted the lowest percentage of total integral citation use in the Published group, with the student groups showing similar levels of coverage. Two items, both discourse act verbs (*argue* and *suggest*) were shared across the three groups. The groups varied in terms of which sharing category represented the highest proportion of top 10 usage: L2S – L2S/Published sharing, L1S – all group sharing, Published– L1S/Published sharing. Overall, when the different combinations of sharing are considered together, L2S writers' usage shared most in common with Published writers, while L1S usage most closely resembled that of the

L2S group. Total sharing was lower overall than in the preceding disciplines, with the highest

percentage (72.59%) found among Published writers, and the lowest (61.05%) in the L2S group.

4.4.2.4(ii)f Physiotherapy

Table 91a-10 most frequently used reporting verbs for each writer group in Physiotherapy

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
claim	3	11.11	suggest	30	21.28	find	12	13.33
according to	2	7.41	find	11	7.80	report	6	6.67
describe	2	7.41	agree	9	6.38	suggest	5	5.56
explore	2	7.41	discuss	8	5.67	conduct	4	4.44
find	2	7.41	describe	7	4.96	analyse	3	3.33
study	2	7.41	support	7	4.96	examine	3	3.33
state	2	7.41	conduct	6	4.26	note	3	3.33
report	2	7.41	present	6	4.26	according to	2	2.22
identify	2	7.41	argue	4	2.84	demonstrate	2	2.22
suggest	1	3.70	confirm	4	2.84	describe	2	2.22
show	1	3.70				observe	2	2.22
point out	1	3.70				recommend	2	2.22
focus on	1	3.70				show	2	2.22
argue	1	3.70						
alert	1	3.70						
advocate	1	3.70						
Total %		96.29	Total %		65.25	Total %		53.31

Table 91b-Top 10 reporting verbs item sharing between writer groups in Physiotherapy

	L2S	L1S	Published
Total frequency / % of total top 10 usage	26/100	92/100	48/100
Shared between all groups frequency / % of total top 10 usage	5/19.23	48/52.17	19/39.58
Shared between L2S/L1S frequency / % of total top 10 usage	1/3.85	4/4.35	n/a
Shared between L2S/Published frequency / % of total top 10 usage	5/19.23	n/a	10/20.83
Shared between L1S/Published frequency / % of total top 10 usage	n/a	6/6.52	4/8.33
Total shared % of total top 10 usage	42.31	63.04	68.75
Unshared frequency / % of total top 10 usage	15/57.69	34/36.96	15/31.25

Although the ten most commonly used reporting verbs accounted for 96.29% of total integral

citation usage in the L2S group, this must be considered in light of the very small size of this

sub-corpus, and the correspondingly low frequencies. Published writers showed the lowest

proportion at 53.31%. Of the three items (*describe, find, suggest*) shared across all groups, two were discourse act verbs and one a research act verb; this was one of only two disciplines in which research acts were more common than discourse acts among Published writers. For L1S and Published writers, all group sharing represented the highest proportion of top 10 usage, while for L2S writers this was equal to L2S/Published sharing. With two and three-way sharing combined, L2S usage shared most with the L1S group, while L1S usage most closely resembled that of the Published group. Highest total sharing was found among Published writers, with lowest in the L2S group.

4.4.2.4(ii)g Social Work

Table 92a-10 most frequently used reporting verbs for each writer group in Social Work

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
suggest	8	10.26	suggest	75	16.23	find	36	6.88
define	7	8.97	state	34	7.36	suggest	29	5.54
highlight	6	7.69	argue	33	7.14	argue	24	4.59
argue	5	6.41	describe	25	5.41	note	23	4.40
describe	4	5.13	according to	23	4.98	according to	19	3.63
develop	4	5.13	find	15	3.25	identify	19	3.63
state	4	5.13	propose	12	2.60	report	17	3.25
according to	3	3.85	affirm	10	2.16	show	11	2.10
find	3	3.85	highlight	9	1.95	describe	10	1.91
explain	3	3.85	note	8	1.73	examine	9	1.72
point out	3	3.85	outline	8	1.73	point out	9	1.72
Total %		64.12	Total %		54.54	Total %		39.37

Table 92b-Top 10 reporting verbs item sharing between writer groups in Social Work

	L2S	L1S	Published
Total frequency / % of total top 10 usage	50/100	252/100	206/100
Shared between all groups frequency / % of total top 10 usage	23/46.00	171/67.86	118/57.28
Shared between L2S/L1S frequency / % of total top 10 usage	10/20.00	43/17.06	n/a
Shared between L2S/Published frequency / % of total top 10 usage	3/6.00	n/a	9/4.37
Shared between L1S/Published frequency / % of total top 10 usage	n/a	8/3.18	23/11.17
Total shared % of total top 10 usage	72.00	88.10	72.82
Unshared frequency / % of total top 10 usage	14/28.00	30/11.90	56/27.18

As in the other disciplines, the proportion of total integral citation use accounted for by the ten most frequently used items was lowest among Published writers. L2S writers showed the highest proportion in this discipline. Five items (*suggest, argue, describe, according to, find*) were shared across all three groups, the majority of these being discourse act verbs, and this sharing category represented the highest proportion of top 10 usage for all three groups. Overall, if the categories of sharing are considered together, both student groups shared more with each other than they did with the Published group. Sharing overall was highest in the L1S group, and lowest in the L2S group.

4.4.2.4(ii)h TESOL

Table 93a-10 most frequently used reporting verbs for each writer group in TESOL

L2S			L1S			Published		
Rep. verb	F	%	Rep. verb	F	%	Rep. verb	F	%
according to	170	13.07	suggest	178	11.01	find	78	8.35
suggest	82	6.30	state	76	4.70	argue	49	5.25
state	72	5.53	according to	75	4.64	describe	36	3.85
argue	54	4.15	find	64	3.96	show	29	3.10
mention	44	3.38	note	55	3.40	suggest	25	2.68
claim	36	2.77	argue	54	3.34	note	22	2.36
find	31	2.38	propose	46	2.84	according to	20	2.14
point out	22	1.69	discuss	37	2.29	report	20	2.14
cite	21	1.61	define	35	2.16	investigate	18	1.93
refer to	20	1.54	describe	35	2.16	point out	18	1.93
Total %		42.42	Total %		40.50	Total %		33.73

Table 93b-Top 10 reporting verbs item sharing between writer groups in TESOL

	L2S	L1S	Published
Total frequency / % of total top 10 usage	552/100	655/100	315/ 100
Shared between all groups frequency / % of total top 10 usage	337/61.05	371/56.64	172/54.60
Shared between L2S/L1S frequency / % of total top 10 usage	72/13.04	76/11.60	n/a
Shared between L2S/Published frequency / % of total top 10 usage	22/3.99	n/a	18/5.71
Shared between L1S/Published frequency / % of total top 10 usage	n/a	90/13.74	58/18.41
Total shared % of total top 10 usage	78.08	81.99	78.73
Unshared frequency / % of total top 10 usage	121/21.92	118/18.02	67/21.27

Again, Published writers showed the lowest proportion of total integral citation usage accounted for by the top ten items, at 33.73%. The four items (*according to, suggest, argue, find* - mostly discourse related) shared across all groups made up the largest proportion of sharing in each case. With two and three-way sharing combined, L2S writers shared more with their L1S counterparts, while the L1S group had more in common with the Published writers. Total sharing was at a similar level in all three groups, indicating consistency between the groups in terms of which reporting verbs are used most commonly.

4.4.2.4(iii) Summary of Within-Disciplines Results

Citations rates per 1000 words were lowest among Published writers in six of the eight disciplines, the exceptions being Fashion, where L2S rates were marginally lower, and Physiotherapy, where Published rates were highest but in which the superscript system may have affected the overall number of citations. Non-integral citations were generally more widely used than integral forms across writer groups and disciplines, although here too there were exceptions – Published writers in Languages, and L1S writers in Physiotherapy and TESOL. The use of non-reporting forms in integral citations was highest proportionally among Published writers in seven of the eight disciplines, with use higher among L2S writers only in Fashion. With the exception of Social Work, where L1S writers showed higher usage, in all disciplines the preposition *according to* was used proportionally more by L2S writers than by either of the other groups. Comparing usage of the various reporting verb types, discourse act verbs predominated in all writer groups in the majority of disciplines. Only Forensic Psychology and Physiotherapy differed, with all three writer groups favouring research act verbs in the former, and Published writers favouring them in the latter. Cognition acts verbs were the least

used of the three types in all but one case – that of L1S writers in Education, who used research act verbs least commonly.

 The proportion of total integral citation use accounted for by the ten most commonly used reporting verbs was lowest among Published writers in every discipline, highest for L2S writers in five disciplines, and highest for L1S writers in three. In terms of item sharing, *suggest* was the only reporting verb to be shared among all writer groups in every single discipline. With only a few exceptions, for all writer groups in all disciplines, the largest proportion of top 10 usage was composed of items shared between all three writer groups. Overall, in four disciplines - Fashion, Forensic Psychology, Social Work, and TESOL, L2S writers' usage was more similar to that of their L1S counterparts than to the Published group. L1S writers themselves shared more with the L2S writers than they did the Published group in five disciplines - Education, Fashion, Forensic Psychology, Languages, and Social Work.

4.5 RQ3 Examination of Pre-sessional Materials

An overview of the observations made in examining the existing pre-sessional materials is provided here. A more detailed account of how this examination complements the main body of results, as presented in the preceding sections, is given within and throughout the discussion, thus helping to create a wider context for the corpus results in terms of how they relate to the existing materials, and the possible implications for academic writing pedagogy.

4.5.1 Vocabulary

AWL vocabulary is common in the existing materials (426 of 570 families represented), occurring not only within examples, tasks, questions and activities, but also as the focus for direct instruction. This includes vocabulary centred on contrast and comparison, cause and

effect, reporting structures, affixes, cohesion, comparatives and superlatives, hedging, and synonyms and antonyms.

4.5.2 Lexical Bundles

With the exception of a limited number of 2- and 3-word bundles used as cohesive devices, and brief references to academic versus non-academic multiword sequences, there is no direct instruction on lexical bundles in the existing materials.

4.5.3 Hedging

Both modals and various other forms of lexical hedge are introduced in the existing materials, including lexical verbs for opinion and speculation, and adverbs and adjectives for certainty/uncertainty. Tasks require students to use, identify, and reformulate examples of lexical hedging.

4.5.4 Citations

Citation is extensively covered in the existing materials. This includes contrasting integral and non-integral forms, presenting a variety of reporting verbs and discussing appropriate usage, and outlining the different referencing strategies available to academic writers.

4.6 RQ4 – Perceptions of Pre-sessional Students

A general overview of the comments expressed by participants in the semi-structured interviews is provided here so as to illustrate the most noticeable trends and commonly expressed opinions. However, participants' comments are considered in greater detail, against the backdrop of the findings from the corpus analysis and examination of existing pre-sessional materials, within the discussion.

4.6.1 Vocabulary

Participants commonly expressed the idea that their Level 7 academic courses required them to know/learn specialised, subject-specific vocabulary, with some commenting that this kind of vocabulary was not taught on the pre-sessional, as students were in mixed-discipline cohorts. Other difficulties relating the academic vocabulary noted by participants included relying on using the same words repeatedly due to a lack of suitable alternatives, and problems distinguishing between academic and non-academic vocabulary.

4.6.2 Lexical Bundles

The majority of the participants commented that lexical bundles, to their recollection, had not featured as a specific teaching focus on the pre-sessional course. Two participants recalled learning some multi-word units related to cohesion, but could not remember having received any other explicit instruction on lexical bundles.

4.6.3 Hedging

Most of the participants recalled receiving instruction on hedging during the pre-sessional course, although their recollections of how detailed this had been and the extent to which the importance of this writing feature had been emphasised varied. All the participants commented that hedging was important when writing for their academic courses, although the most commonly noted problem was feeling that they lacked sufficiently varied means of hedging statements in academic writing.

4.6.4 Citations

Participants varied in terms of both their confidence in citation use, and their opinions of how effectively it had been taught on their pre-sessional courses. While some felt that citation use had been covered effectively and extensively, and as a consequence felt confident about their

own citation practices, others commented that they still felt uncertain about the right and wrong ways to write citations. One even responded that the pre-sessional instruction had left them confused, as they had been told different things by different tutors. In terms of varying citation forms, most respondents commented that they either favour one form most of the time, or vary their use simply for the sake of variety, with only one participant commenting on the rhetorical aspects of different citations forms.

5 Discussion

5.1 Overview of Chapter

This chapter will draw together the results from the corpus analysis, interviews, and examination of pre-sessional materials in order to answer the research questions. These concern the four academic writing features in terms of the extent of disciplinary variation, the differences between the writer groups, the representation in existing pre-sessional materials, and the perceptions of past pre-sessional students. Each of the four writing features will be discussed in turn, firstly with regard to the across-disciplines results, and secondly in terms of the results of the within-disciplines analyses. Possible explanations for the study's findings will be offered, and their significance will be examined with reference to the existing literature. Pedagogical implications in relation to each of the four features will then be discussed, with a view to informing EAP pre-sessional writing instruction, and contributing to the existing knowledge in these areas. This will be of benefit in helping to optimise teaching outcomes for future pre-sessional participants, given the difficulties often experienced by L2 students when it comes to writing academically in English. As the interviewees commented:

'I'm still struggling to make it sound professional.' (P1)

'Actually my biggest weakness is writing…If I could have a class I would like to do more, but I don't know how to improve.' (P2)

'Definitely more complex…it's a very specific subject but [although] I have some knowledge that makes it less difficult, compared with pre-sessional course, it's more complex.' (P4).

5.2 Vocabulary

5.2.1 Variation across Disciplines

The level of AWL coverage for the corpus as a whole (11.93% of tokens) was slightly higher than that reported in some previous studies of academic corpora, including Chen and Ge (2007) (10.07%), Coxhead (2000) (10%), and Hyland and Tse (2007) (10.6%). One explanation for this may be the lack of science disciplines in this corpus, as AWL coverage for science has been shown to be lower than for other disciplines such as arts, commerce, law, or social sciences (Coxhead, 2000; Hyland & Tse, 2007). A corpus with no hard sciences might therefore be expected to have higher overall coverage.

In terms of RQ1, the disciplines did not show tremendous variation in terms of AWL coverage among any of the writer groups, although variation was largest among L2S writers (5.29% between the highest and lowest) and smallest in the Published group (2.32% between the highest and lowest). Why this might be is hard to say definitively, but it is possible that text type and/or topic may play an influencing role. All texts from the Published writers were research articles, and the topics covered (beyond the obvious disciplinary constraints) were not restricted. Among the students however, there were various text types, and perhaps more importantly, a very limited range of topics within each discipline, since some sub-corpora consisted entirely of only one or two assignment tasks. It is not unreasonable to imagine that the difference in vocabulary use between corpora in two given disciplines might be more pronounced if the majority of texts in each focus on a specific topic (therefore 'concentrating' repeated uses of vocabulary specific to that topic). Conversely, if each discipline contains a wide variety of topics, this topic-focused vocabulary use will be less noticeable, and a more general selection of academic vocabulary may be used, thus allowing for a greater degree of

commonality between the disciplines. This may be why both student sub-corpora showed more disciplinary variation in AWL coverage than did the Published sub-corpus. The effect of limited topic range on AWL coverage is also noted by Wang (2014).

In terms of the disciplines themselves, the writer groups showed limited commonality in terms of 1K and 2K, AWL, or Off-list type coverage. For both L2S and Published writers the lowest AWL type coverage was found in Fashion, the lowest 1K and 2K coverage occurred in Social Work, and the lowest Off-List coverage was in Languages, but the patterns of disciplinary variation were not consistent among writer groups for the most part. Why the most similarity in type coverage across disciplines should be between L2S and Published writers rather than, as might be expected, between L2S and L1S (given the shared status as 'novice' academic writers), or L1S and Published writers, is not immediately obvious, but as discussed below in Section 5.2.2, strategies used by L2S writers in lexical decision making may have an influence.

Looking at the extent of item sharing across the disciplines, there is little consistency to be found in terms of which disciplines showed more or less sharing in different writer groups. Examining which writer groups showed more or less sharing in different disciplines however, one pattern does emerge – in six of the eight disciplines, it was the Published writers who showed the highest proportion of item sharing across the disciplines. This suggests that disciplinary variation in academic vocabulary (or at least, academic vocabulary as represented by common use of AWL families) may be more pronounced among the student writers in this study. Once again, the limited range of topics in each discipline may contribute significantly to this increased variation. Nonetheless, for all writer groups in the across disciplines analysis, disciplines included unshared words that could potentially be seen as more relevant to one subject area than another, and as noted in Section 4.2.2.5(iii), the same could be said of some

items common to the top 20s of all writers groups *within* certain disciplines. This would seem to indicate that where the AWL is concerned, certain items are more commonly found in some disciplines than in others. Indeed the occurrence of AWL families in different subject areas is known to differ (Coxhead, 2000). It is also important to remember that even though a given family may be shared between different disciplines, the possibility exists for the same words to be used with differing meanings in different disciplinary discourse communities (Hyland & Tse, 2007), such that variation may be greater than is indicated simply by item sharing. It is clear from previous studies (Durrant, 2014; Hyland & Tse, 2007, 2009; Vongpumivitch et al., 2009) that academic vocabulary is very much a product of disciplines. Importantly, the results here also suggest that, even when it comes to the most frequently used items from an apparently generic list such as the AWL, different disciplines do, at least to some extent, employ academic vocabulary differently.

5.2.2 Variation within Disciplines

In six of the eight disciplines, the proportion of total AWL family tokens represented by the 20 most commonly used families was lowest among Published writers when compared to the L2S and L1S writers. With reference to RQ2, this indicates that the student writers may commonly rely on a narrower variety of AWL families than the Published writers. This, as has been suggested in the case of lexical bundles, may be a result of novice writers attempting to avoid erroneous use by sticking to the familiar (Hyland, 2008b). When overall AWL type coverage is considered, this was highest among Published writers in three disciplines, and highest among L2S writers in five, while L1S writers showed the lowest AWL coverage in seven of the eight disciplines (although differences in coverage between student writers were minimal in some cases). One possible explanation for the higher rates of AWL type coverage among L2S writers

compared to their L1S counterparts may be an increased awareness of and effort to use academic vocabulary on the part of students for whom English is not a first language. Being conscious of the fact that they may not have as extensive a lexical resource to draw upon, it is conceivable that the L2S writers pay more attention when it comes to attempting to use vocabulary that demonstrates appropriate register or familiarity with the subject area. L2 student writers have been shown to use a complex variety of decisions when it comes to their use of lexis, and to employ strategies such as discussing vocabulary use with other L2 students, checking meanings in dictionaries, and utilising direct quotes in order to include specific 'academic' words (Coxhead, 2012). It is of course, entirely possible that L1 student writers may also adopt such strategies when writing assignments. Nonetheless, the results here tend to suggest that, at least where AWL families are concerned, L2S writers show a type coverage more similar to that of the Published writers. Given their 'expert' status, these published academic writers could reasonably be expected to incorporate academic vocabulary at levels appropriate to the register, and to their academic discourse communities. For L2S writers to achieve a generally closer approximation to this than their L1 counterparts when it comes to AWL type coverage, is perhaps encouraging from the point of view of EAP vocabulary pedagogy.

One of the most significant findings in terms of vocabulary usage is the fact that total item sharing was lowest among Published writers in every discipline, with the proportion of top 20 usage shared being substantially lower than that for either student group. The student writers themselves showed levels of total item sharing that were relatively consistent with each other, and the results indicate that in all the disciplines, the most frequent AWL families used bore more similarity between the student groups than between either student group and the Published writers. As discussed in the preceding section, topic may have influenced this. The

corpora for both student groups in each discipline were composed of responses to the same assignment rubrics, and it is therefore unsurprising perhaps that L1S and L2S writers employed similar lexical resources through which to tackle them. Published writers on the other hand, were not restrained in such a way, given that the 45 research articles for each discipline could potentially involve 45 different topics (albeit in a related field). This represents far less of a potential constraint when it comes to appropriate lexical choices, and we might therefore expect item sharing to be lower among the Published writers.

5.2.3 Pedagogical Implications

Academic vocabulary has been a focus of EAP instruction for some considerable time, although the notion of a general academic vocabulary is a contentious one (Hyland & Tse, 2007). While many efforts have been made to compile lists of academic words (Browne et al., 2013; Coxhead, 2000; Gardner & Davies, 2014 being some notable examples), the importance of discipline is widely recognised. This creates issues for how best to approach EAP vocabulary pedagogy in contexts such as pre-sessional courses. The difficulty is perhaps best summarised by Durrant (2016), who argues that while academic vocabulary may be largely discipline specific, such teaching is not always practical in mixed-discipline cohorts, and as such, generic vocabulary lists are potentially very useful. He goes on to note however, that such generic vocabulary may in fact be very limited, and that therefore discipline-specific lists may be a better use of students' time where possible.

An indication of how the issue of academic vocabulary may affect EAP learners can be gleaned from the interview participants, when asked about the extent to which they felt the pre-sessional course had prepared them for the vocabulary they require in their academic courses:

'The pre-sessional was mixed people with different backgrounds and people that were going to study different fields. So now I'm in my specific field it's more technical and new technical words.' (P4)

'Because in pre-sessional it's every different, students from different subjects, so it's impossible to learn, like to focus on your subject or vocabulary.' (P5)

'I think they should add more specific words for our majors. Architecture has lots of words, actually all majors have some complex words.' (P2)

'...sometimes the papers and topics were related to our subject. I think architecture vocabulary is easy. It's not something like law, so I don't think it's need to study architecture vocabulary.' (P6)

Academic vocabulary clearly runs throughout the existing pre-sessional content, with the in-house materials alone containing 426 of the 570 AWL families. In addition to the general occurrence of a wide variety of academic vocabulary in questions, tasks and activities; example sentences, paragraphs and essays; and sections presenting information, the materials also contain direct instruction related to vocabulary. Oxford EAP includes a vocabulary focus as part of each unit, and these deal with topics such as collocations, evaluative adjectives, affixes, reporting structures, verbs for essays, cohesive language, comparison and contrast, cause and effect, argument, and recognising and using general, academic, and technical vocabulary. Many of these topics also feature as part of the in-house materials, in addition to others such as comparatives and superlatives, vocabulary for hedging and boosting, and synonyms and antonyms. Learners are exposed to a broad range of academic vocabulary throughout these materials, and given numerous opportunities to encounter vocabulary and practise using it,

not only in the tasks directly focusing on vocabulary itself, but also as they complete other sections of the materials. Nonetheless, a number of the interviewees commented that academic vocabulary still presents them with problems. Some are unclear as to which words are academic and suitable to use in assignments, while P1 noted that lacking alternatives, he tends to re-use the same words repeatedly. P2 observed that her lecturers use technical or professional words that she doesn't understand, and P3 said that her main problem in writing was vocabulary, as she has a limited range and feels that most of the words she knows are 'not academic enough for essays'.

L2 writers in this study actually showed the highest AWL type coverage of the three writer groups in five of the disciplines, so if there is an issue, it is not with the proportion of different words in their texts that are academic, assuming these words are being used correctly. At the same time, in all but one discipline, L2 writers employed fewer AWL families than Published writers (although L1 writers employed fewer still in six of the disciplines), and in six disciplines their 20 most commonly used AWL families accounted for a higher proportion of AWL tokens than was the case for Published writers. It may be then that rather than further encouraging learners to use academic vocabulary, of more importance is encouraging them to use a wider variety, and to not rely restrictively on those words they are most familiar with.

It is obvious that the breadth and quantity of new vocabulary that might be of use to L2 students far exceeds the time and resources available in a pre-sessional course. Academic vocabulary already occupies a prominent place in the existing materials, and the amount of vocabulary that learners can absorb for productive use in a given lesson is limited, especially if there is a high level of difficulty (McCarten, 2007). Perhaps therefore, rather than encouraging more varied use among learners by attempting simply to expose them to more vocabulary on

the course, it may be beneficial to devote time to raising learners' awareness of strategies they themselves can employ *after* the pre-sessional, when they will also be exposed to the discipline-specific vocabulary of their own subject areas. A number of the interviewees commented on strategies they use to learn vocabulary, or to improve how they make use of it in assignments. These include asking others to explain unfamiliar words that arise in conversation (P1), writing down unfamiliar words from books or journals and checking them on the internet (P2) or in a dictionary (P3, P6), using associations and collocations (P3), discussing vocabulary with L1 classmates (P4), and recording lectures and using a vocabulary notebook (P5). Pre-sessional learners could be made aware of a variety of effective learning strategies, and of the importance of continuing to learn, revisit, and make productive use of new vocabulary. Understanding why they use a given strategy, selecting those that complement each other, combining both cognitive and metacognitive strategies, managing strategy use, reviewing and practising target words, and being conscious of their learning are all features of effective L2 learners (Mokhtar et al., 2017; Oxford, 2002), and higher levels of vocabulary learning strategy awareness are also a way of increasing learner autonomy and independence (Nosratinia & Zaker, 2015). The existing materials already include sections on reading and notetaking strategies, so the addition of vocabulary learning strategy instruction would not be out of place. This could include ideas such as guessing from context, dictionary training, keeping learning journals, using research tools, and guidance for effective vocabulary notebooks, with entries such as parts of speech, pronunciation, collocations, and example sentences, rather than simply an L1 translation (Diaz, 2015; McCarten, 2007; Mokhtar et al., 2017; Schmitt & Schmitt, 1995). Additionally, instruments to assess and raise awareness of vocabulary learning strategy use could be used, such as questionnaires like Gu & Johnson's

(1996) Vocabulary Learning Questionnaire, or Mizumoto & Takeuchi's (2008) Vocabulary Learning Strategy questionnaire.

In terms of how the pre-sessional might address disciplinary variation in academic vocabulary, a pragmatic approach is probably the most realistic. Learners may find vocabulary specific to their own academic discipline very helpful, but such focused instruction may not be practical in this context. As Durrant (2014) observes, while discipline-specific word lists are a frequently explored option, approaching from a more general view of how academic vocabulary varies across texts and disciplines may be more appropriate in multi-disciplinary groups. Existing pre-sessional materials already incorporate example texts and sentences from a variety of disciplines, so perhaps such materials could be used to highlight academic vocabulary. Instructors could encourage any learners for whom a given text is subject related, to pay particular attention to the vocabulary – for example to highlight unfamiliar words and engage their vocabulary learning strategies to explore them further in terms of meaning, collocations, word forms, synonyms, and so on. It may even be feasible to have a variety of potential materials derived from a wider range of disciplines, and to have different learners engage with different texts, so that individuals can gain greater exposure to the kind of vocabulary they may encounter in their studies. Contextual environments reflecting varying disciplinary norms are important if we are to teach the most useful and relevant vocabulary to students (Hyland & Tse, 2007). Such subject-differentiated materials, combined with productive tasks based on individual learners' subject areas, could constitute a valuable opportunity to encounter, explore, revisit and re-use new words, encompassing Nation's (2007) four strands of balanced vocabulary instruction – meaning focused input and output, deliberate vocabulary learning, and developing fluency. It could also be of benefit to instructors, giving them the opportunity

to develop skills in analysing texts from different disciplines and identifying words that may be particularly useful to learners (Coxhead & Byrd, 2007). Additionally, as mentioned above, encouraging independent learning and autonomy when it comes to vocabulary would likely benefit students far beyond completion of the pre-sessional course. Approaches such as those laid out above may help to encourage more varied vocabulary use by L2 students, as well as addressing to some extent the issue of disciplinary variation in academic vocabulary. By enhancing the focus on learners' own independent strategy use, the burden on EAP instructors when it comes to academic vocabulary need not be unduly increased.

5.3 Lexical Bundles

5.3.1 Variation across Disciplines

The results for the most common 4-word bundles in the corpus as a whole show that this corpus bears some similarity to corpora examined previously. Bundles such as *it is important to, on the other hand, in the context of, as a result of, as well as the, in the case of, at the same time, are more likely to, in the form of*, *the extent to which* and *in terms of the,* which were among the 20 most common bundles in this corpus, have been shown to occur with high frequency in other corpora of academic writing (Ang & Tan, 2018; Hyland, 2008a; Hyland & Jiang, 2018; Simpson-Vlach & Ellis, 2010). Disciplinary variation in lexical bundle use is also a common feature of previous studies (see for example, Cortes, 2004; Hyland, 2008a, Liu, 2012) and in this study too, there was evidence of this. With reference to RQ1, rates of overall item sharing across disciplines among the 20 most commonly used bundles were generally highest for the Published writers, which suggests that disciplinary variation in student usage may be more pronounced. Indeed, for both L2S and L1S writers, six of the eight disciplines shared less

than 30% of their top twenty usage overall, while among Published writers, only one discipline exhibited overall sharing below 30%.

Examining this variation more closely, L2S writers in the various disciplines shared 12 different bundles in total, ten of which were among the 20 most common bundles in the corpus as a whole. That many of these shared items have also, as mentioned above, been identified as common in academic corpora before, suggests that they may be regarded as among those bundles that 'help identify a text as belonging to an academic register' (Hyland, 2008b:42), and might therefore be expected to occur frequently across a range of disciplines. Among the L1S writers, widespread sharing across disciplines was more limited, but of the 15 items shared, nine were among the 20 most common in the corpus, again pointing to some use by this writer group of the very common, subject non-specific bundles associated with academic register. Published writers showed a generally higher level of sharing and thus uniformity across disciplines. Of the 30 shared bundles, 15 represented items within the top 20 for the whole corpus, again including many of the bundles found commonly in previous examinations of academic corpora. It is perhaps unsurprising that these 'expert' writers, being experienced in the academic discourse community, would make use of a wide variety of bundles common in academic register.

As noted however, for the majority of disciplines in the L2S and L1S sub-corpora, unshared bundles made up the greater proportion of top 20 usage. High unshared usage across disciplines suggests disciplinary variation; however, it is important to look at the unshared items themselves, since any variation may have a variety of underlying causes. These may include bundle use that is inappropriate to academic register (Ädel & Erman, 2012) and thus anomalous, or effects arising as a result of specific assignment topics rather than from the

disciplines themselves (Chen and Baker, 2010). This second factor may play a particularly important role given that all disciplines in both student corpora included multiple texts submitted for the same assignment rubrics. The fact that the most frequently occurring bundle in The L2S Forensic Psychology sub-corpus was *that Mr. Taylor has*, referring to the subject of a case study assignment, illustrates this point. Lexical bundles have also been observed to show sensitivity to text type (Durrant, 2017). Finally, a number of the student (L2S in particular) sub-corpora were also relatively small in size, which may amplify the overall influence of individual writers, and therefore limit the inferences that can be made.

In an attempt to identify bundles that may actually represent usage specific to discipline, the potential subject specificity of unshared items was examined. Disciplines showing relatively higher proportions of bundles that can be seen to be subject specific included Languages (for both student groups), Education (For L2S and Published writers), Social Work among the Published group, and Physiotherapy for L1S writers. Conversely, Fashion, TESOL, and Forensic Psychology exhibited lowest potential subject-specificity for L2S, L1S, and Published writers respectively. For these disciplines with less evidence of subject specificity in common bundle use, it is particularly problematic to suggest that any unshared items may reflect bundle use directly related to discipline, since without a far more detailed examination of the context in which these bundles occur, the fact that they are present in one discipline but not another may be indicative of other factors. Indeed, even where more evidence of potential subject specificity is present, the occurrence of unshared bundles as a result of topic or text type cannot be ruled out in the student sub-corpora. It may be safer to suggest disciplinary influence where Published writers are concerned, given that text types were uniform, and topics were not restricted in the same way as in the student sub-corpora. Nonetheless, Hyland (2008a:20)

252

in his examination of research articles, dissertations and theses, does contend that the distribution of 4-word bundles can 'help characterise disciplinary discourses'.

Aside from variation in the specific bundles used, disciplines also show differences in terms of the rates at which bundles are used. Among L2S writers, higher rates of bundle use were seen in Education and Languages, while for L1S students, Languages and Social work exhibited the highest rates. As with bundle sharing, there was more uniformity across disciplines for Published writers, with rates differing to a much smaller degree and being more consistent with levels observed in studies such as Conrad and Biber (2005). That variation in the extent to which bundles are used in different disciplines is found here, mirrors the findings of previous studies, including Kwary et al. (2017), who examined research articles in various science disciplines, and Hyland (2008a) in his study comparing engineering, biology, business studies, and linguistics. However, the fact that the variation in this study is less pronounced among the Published writers, suggests that students' levels of usage may be influenced by factors other than disciplinary convention. It is hard to say what these factors may be, although as with bundle sharing, it is possible that assignment topic may play a role.

While the reason for differences arising cannot be precisely determined, it is true to say that for the writers in this study, bundle use across disciplines is not uniform, although Published writers do show a general tendency towards less variance than their student counterparts, perhaps as a result of a greater familiarity with the general conventions of academic discourse, or perhaps because of influences like text type and topic. To different degrees in different disciplines, these writers are employing 'different resources to develop their arguments, establish their credibility and persuade their readers' (Hyland, 2008a:20).

5.3.2 Variation within Disciplines

The extent to which the different writer groups shared lexical bundle use differed depending on discipline, but there were noticeable patterns to be observed. Certain bundles were common to all three writer groups in a variety of the disciplines, notably *it is important to, on the other hand,* and *the end of the*, suggesting that there may be some 'core' bundles, which transcend disciplinary boundaries and are commonly employed by student and published academic writers alike. This would generally support the view of Simpson-Vlach & Ellis (2010) in their formulation of the AFL, that such core bundles do exist. However, the number of these bundles was small, and it could equally be argued that, as in Hyland's (2008a) study, in which only four bundles were found to be common to all of the top 30 bundles lists in four disciplines, this is further evidence of variation by discipline. It should be noted that more of such bundles may have been identified in this study had the focus on bundle sharing not been restricted to 4-word bundles, and only the 20 most commonly used of these, since in most cases the cut-offs of 40 times per million words and three texts resulted in bundle lists longer than 20 items.

Despite some commonality however, in terms of RQ2, there were also differences in bundle usage between the writer groups. The total proportion of top 20 usage shared tended to be highest among the L2S writers in the majority of disciplines. This suggests that these writers are relying more heavily on a limited selection of commonly used bundles, whereas L1S and Published writers draw on a wider variety in their frequent usage - something that has been argued previously (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004). This may be because L2 writers have a tendency to stick to using expressions they are comfortable with (Paquot, 2007). In bundle usage the student writer groups generally showed more commonality with each other than they did with the published writers, who also, in the

majority of disciplines, showed the lowest level of item sharing overall. This echoes the findings of previous studies in which bundle usage between students and 'expert' writers has been shown to differ. Published writers have been shown to make frequent use of many bundles that are only very rarely used by student writers, who may avoid using unfamiliar bundles for fear of erroneous use (Cortes, 2004; Hyland, 2008b). Chen & Baker (2010) reported similarity between L1 and L2 student writers' usage of bundles, as well as divergence between student and published writers. Looking at the overall rates of bundle usage, Published writers in this study, in all but one of the disciplines, employed bundles less frequently overall than either of the student groups, something also found by Hyland (2008b). There may be a number of explanations for these differences between student and published writers. It is possible that, as Hyland observes, novice writers may rely more heavily on formulaicity, but it could equally be the case that genre/text type plays a role. The research articles in this study reported on practical research, while the student assignments included a variety of tasks such as case studies and critical analyses, as well as research reports. Bundles found in the top 20 among Published writers but not either of the student groups included *is positively related to, the results of the, the purpose of this, the extent to which,* and *in the present study*, and bundles such as these may reflect the more prominent research focus of the Published writers.

5.3.3 Pedagogical Implications

The results of the corpus analysis suggest that there are, as previously proposed, a number of what might be termed 'core' bundles, which occur with relatively higher frequency across disciplines. Such bundles could form the basis of a useful tool in EAP writing instruction (Shahriari, 2017), since it is important for L2 learners to use multiword sequences accurately and appropriately (Bychkovska & Lee, 2017). Of particular importance for pedagogy, are the

greater apparent reliance on the most commonly used bundles among L2S writers in this study, and the lower levels of item sharing for Published writers in the majority of disciplines. These also indicate that student writers (both L2 *and* L1 in fact) may benefit from an increased awareness of these sequences. However, Chen & Baker (2010) question the extent to which lexical bundles identified through CL techniques have appeared in curricula and materials. Focusing specifically on the pre-sessional course included in this study, similar doubts could be expressed. Most of the interviewees, when asked, could not recall being given any specific instruction on lexical bundles for academic writing, while two commented on receiving limited teaching, but only for the purposes of cohesion – 'Well, yes we have talked about using linking words in the pre-sessional class, but just a few examples' (P5), and 'Yes, but only for cohesive stuff like *as a consequence*, or *as a result*, or things to use between two sentences, but for the specific things in academic writing, I don't think I received teaching on that' (P3).

The examination of the existing pre-sessional writing materials would seem to confirm that these recollections are accurate; no specific or explicit focus on lexical bundles is currently included. A small number of 2 and 3-word bundles such as *as a result* and *in conclusion* are referenced in activities relating to cohesive devices, and a unit on academic style provides a few examples of multiword sequences that may be appropriate alternatives to less academic lexical choices (*it seems that* as opposed to *I think*, or *an increasing number of* instead of *more and more* for example), but lexical bundles as a specific focus in writing instruction are absent.

If then, as has been suggested in numerous previous studies (Ang & Tan, 2018; Cortes, 2004; Hyland, 2008a, for example) and as the results here would seem to suggest, bundles are an important part of academic writing and would be a beneficial inclusion in EAP instruction, what might this look like in a pre-sessional context? It could take the form of awareness-raising

activities, with instructors drawing attention to bundles in class materials (Bychkovska & Lee, 2017; Byrd & Coxhead, 2010; Liu, 2012) in order that learners notice the discourse functions of different bundles and the contexts in which they occur (Cortes, 2004). This could be followed by opportunities to manipulate, use, and produce bundles in writing (Hyland, 2008a). Cortes also suggests introducing bundles with functions such as organising text, in conjunction perhaps with discourse markers – moving from single words to longer combinations. The results of the corpus analysis do suggest however that bundles show variation across the disciplines, and although this is less pronounced among the published writers, it is important to remember that had the analysis included disciplines such as hard sciences, engineering, or mathematics, the extent of disciplinary variation, even among the expert writers, may have been much greater. It would therefore seem that if bundles are to be taught, the ideal model would involve some form of disciplinary differentiation. However, as Ang & Tan (2018) observe, such disciplinary specificity can be difficult in EAP classrooms. In a pre-sessional context where learners are not divided on the basis of their study discipline, it may not be possible to tailor teaching for each learner's particular subject area. Nonetheless, there are possibilities. It may be beneficial, for example, to contrast 'core' bundles, with those more specific to disciplines (Pérez-Llantada, 2014), thus furnishing learners with not only a basic lexicon of the most common bundles, but also some insight into how sequences might be used in different subject areas. The existing pre-sessional materials already include example essays, articles, and activities based around a wide range of subjects including consumer behaviour, economics, environmental science, technology, education, and social issues. Such materials could be used to draw attention to lexical bundles (both commonality and contrast) in various disciplinary areas such as humanities, hard sciences, arts, and business. While this would not give every

learner a precise insight into bundle use in their own specific academic field, it would at least serve to highlight these features and increase learners' awareness of them and the fact that different academic disciplines may have different conventions. Ideally, materials throughout the course, including those primarily focused on other skills, such as reading, would allow learners to revisit the bundles multiple times. Repeated encounters with authentic, contextualised examples would aid learning, given that building vocabulary knowledge is an incremental process (Byrd & Coxhead, 2010; Pérez-Llantada, 2014). Nonetheless, the question remains – how do we select which bundles should be taught?

 It is clearly important to prioritize the appropriate bundles for teaching (Ellis et al., 2008) and while corpora can improve our understanding of language use and provide realistic models for learners (Hyland, 2008b), we return to the perennial debate over what represents a suitable target text for L2 student writers. The expertise and familiarity with academic writing conventions found in published texts such as research articles clearly contain some value, but as the results of this and numerous studies in the past have illustrated (Gardner & Nesi, 2012; Hyland, 2008b; Jalilifar, 2012a), there are also good reasons to suspect that research articles differ notably from university assignments, not least because of their different communicative purpose and intended audience. Pre-sessional teaching materials should help learners to write the kinds of texts they will need to write, and arguably a corpus-driven approach to materials development should not be based entirely on a genre many students may never go on to tackle. Bychkovska & Lee (2017) argue that an appropriate model for L2 undergraduates is the writing of high-scoring senior undergraduates, while Hyland (2008a) suggests materials be based not only on the kinds of texts students will need to write, but also on those they will need to read. Given the useful insights into bundles gained even in the present, relatively small study of a

limited number of disciplines, I would contend that if we are to develop corpus-driven teaching materials for pre-sessional courses, the target texts for analysis should come both from student, and expert sources. This should also include a variety of disciplines, so as to provide a broad representation of bundle use in academic discourse. Such an approach would allow for the creation of materials that might, as Hyland (2008a) observes, help learners to gain some understanding of bundle use in their own (albeit broad) academic fields. It would also aid in highlighting those bundles that are common across disciplines and academic genres, and might thus be considered to represent more 'universal' core items, of benefit to all learners, irrespective of their subsequent academic courses.

 In terms of selecting the bundles themselves, frequency and range of occurrence are obviously a means of identifying bundles worth teaching, particularly those most common. This study identified bundles such *as it is important to, on the other hand, in the context of, as a result of, as well as the, at the same time*, and *are more likely to* as being particularly frequent, and many of these have, as noted in Section 5.3.1, been similarly highlighted in previous studies, which would suggest they might have some pedagogical value to L2 student writers. Further discipline-specific analyses of academic corpora would help to identify those bundles that may be less widespread but of particular importance in individual subject areas. However, frequency and range may not be the only important factors. It may be beneficial to consider principles like teachability and learnability (Byrd & Coxhead, 2010) for example. Simpson-Vlach & Ellis' (2010:508) Academic Formulas List also included, in addition to frequency, a measure of mutual information (MI), which seeks to quantify the likelihood of words occurring together. They found that bundles scoring highly for both MI and frequency, were regarded by

experienced EAP instructors to have 'more clearly defined functions, and to be more worthy of instruction'.

While future corpus analyses of the nature described above would prove extremely useful in informing teaching content for bundles in EAP contexts, existing studies such as this one and others cited above could represent a good starting point from which to identify those lexical bundles that might be of value to L2 student writers. It may even be possible to modify existing pre-sessional writing materials in the short term so as to highlight these where present and help to raise learners' awareness of these important features. This would seem a valuable approach, given the minimal attention currently given to lexical bundles in this specific context.

5.4 Hedging

5.4.1 Variation across Disciplines

Despite the corpus containing 88 of the 91 lexical hedges from the reference list, over 65% of all hedge usage was accounted for by only the ten most commonly used hedges. This would suggest a considerable uniformity in terms of hedge choices within the corpus as a whole, but is this applicable across disciplines?

The overall highest and lowest average rates of hedging were seen in Forensic Psychology and Fashion respectively. To my knowledge, no previous research comparing exactly the same disciplines as those in this study exists, but there is certainly a precedent for variation in the rates of hedging use across different subject areas. Poos and Simpson (2002) found higher rates in social sciences than in humanities or hard sciences, and Takimoto (2015) noted that rates in marketing (a core component of the Fashion discipline in this study) were lower than those in disciplines such as philosophy and linguistics. Marketing and business studies have

been shown to have higher rates when compared to hard sciences however (Hyland, 2004; Vasquez & Giner, 2008). In the present study, these differences across disciplines were not entirely consistent when the results for the different writer groups were compared. Fashion had the lowest rate for both student groups, but for Published writers, the lowest rates were in Physiotherapy (the closest discipline to a hard science in this study). For both L1S and Published writers, Forensic Psychology was highest, while for L2S writers this was Social Work. While it may be the case that the variation in rates across disciplines observed here is directly related to discipline, other factors such as text type or topic cannot be ruled out. It is entirely possible that certain assignment tasks may require a greater emphasis on speculation, or focus more heavily on the potential implications of primary data, and thus tend towards encouraging more frequent use of hedging.

In terms of the hedging items themselves and the extent to which usage differed across disciplines, while there was some variation, levels of item sharing were generally high (over 80% of top ten usage shared for every writer group and every discipline). This was most noticeable among Published writers. Verbs, and particularly modals, were the most common choices of lexical hedge across disciplines, with *may* being the most common hedge in 17 of the 24 sub-corpora (L2S - four of eight disciplines, L1S – five of eight, Published– eight of eight). Demir (2018) also found *may* to be the most common modal hedge among L1 ELT RA authors. Indeed only one of the 24 sub-corpora – L1S Forensic Psychology – did not have a modal as the most commonly used lexical hedge. This would seem to concur with the findings of studies such as that by Vandenhoek (2018), which found modals to be among the most common epistemic markers in a corpus of assignments by L1/L2 English university students in Ireland. It also reflects Hyland's (1994:247-248) assertion that 'modals are the most easily identified

and widely used means of hedging in academic writing...after modal verbs, the most common means of expressing epistemic modality in written discourse is through the use of lexical verbs'. In six of the eight disciplines the percentage of top ten usage accounted for by modals was higher among L2S writers than in the L1S group. This may suggest, as previously postulated (Hyland & Milton, 1997), a greater preference for modal verbs among L2S writers. The percentage of top ten usage attributable to modals varied only slightly across disciplines for Published writers (lowest 41.27%, highest 46.79%), but much more noticeably in the student groups (L2S – lowest 38.10%, highest 60.67%; L1S – lowest 36.19%, highest 54.49%). If disciplinary conventions in discourse communities are, as has been suggested, well represented in research articles (Hyland, 199b, 2002b; Jalilifar, 2012b; Pérez-Llantada, 2014) then this suggests a degree of uniformity across disciplines not reflected in student usage.

Examining sharing of top ten items overall, L2S writers showed the lowest commonality in five of the eight disciplines, while Published writers demonstrated 100% sharing in five of the eight. This again hints at a limited variation in disciplinary conventions in research articles where lexical hedges are concerned – a uniformity adhered to less consistently by student writers. The L2S/L1S groups did show commonality with each other, in that both groups demonstrated high sharing in Advertising and Forensic Psychology, and exhibited their lowest sharing in Fashion, but whether this reflects any genuine disciplinary convention is unclear.

Item sharing and rates of modal use and overall lexical hedge use in this study paint a picture of greater disciplinary uniformity among Published writers. It may therefore be that where hedging is concerned, as Vold (2006) argues, the influence of discipline is relatively unimportant, and that other factors, such as text type, topic, or even L1, play a greater role in creating variation in usage. However, it must also be remembered that the corpus in this study

included no hard science disciplines, and that previous research, as mentioned above, tends to indicate more noticeable differences between these and disciplines such as social sciences and humanities. It is conceivable that clear disciplinary variation in the use of lexical hedges *is* to be found in research articles, but that such differences were not highlighted by the disciplines included in this study.

5.4.2 Variation within Disciplines

In five of the eight disciplines, Published writers showed higher rates of hedge use than either of the student groups. Among students themselves, hedges per 1000 words were higher for L1S writers than for L2S writers in six of the eight disciplines. These results both converge and diverge from those of previous studies. Yüksel & Kavanoz (2018:110), in an analysis of novice L1 and L2 English, and expert L1 English academic writing, found, as in this study, higher rates of hedging among expert writers. They also noted that 'novice non-native learners opt for expressing their commitment to their propositions more', which would seem to align with the generally higher rates of hedging found among L1S writers in this study. L1 students have been shown to use more hedges than L2 writers (Ädel & Erman, 2012), but have elsewhere been found to use hedges at comparable rates to L2 students (Vandenhoek, 2018). It must be remembered however, that there may be other factors at play. Language, cultural background, and topic can affect the extent to which hedging is employed (Gabrielatos & McEnery, 2005; Hinkel, 2005, 2009), so different cohorts of L2 student writers with different backgrounds may produce varying results. Not reflecting the current findings, novice research reports in dentistry were found to have higher frequencies of hedging devices than those written by experts (Crosthwaite et al., 2017). It is perhaps interesting to note that in the present study, Physiotherapy was one of only two disciplines in which Published writers showed the lowest

rates of hedging. With its healthcare context, this is, of all those in this study, the discipline most closely comparable to dentistry, so the possibility of specific disciplinary influences must be considered. Nonetheless, a general tendency for higher rates of hedging among published academic writers is hinted at by Aull et al. (2017), who cite a number of studies indicating that students demonstrate more limited use of hedges when compared to published academic writers. Part of the reason for this may be the kind of texts students are producing; if assignments are not based on primary data, there may simply be less necessity to hedge (Vandenhoek, 2018).

Looking at the ten most commonly used lexical hedges, a mixed picture can be seen in terms of the proportion of total hedge usage made up by these most common items. In seven of the eight disciplines, it was Published writers who showed the lowest proportion, suggesting that they tend to make use of a less restrictive range of lexical hedges. The highest proportions were seen among L2S writers in four disciplines, and L1S writers in four. Published writers only reached a proportion above 75% in one discipline (Physiotherapy), while for student writers, proportions above 75% were seen more commonly (L1S – four disciplines, L2S – two disciplines). These results are less clear cut than might be expected, given previous findings, which showed a more restricted range of hedges among L2 students (Hyland & Milton, 1997; Vandenhoek, 2018). Again however, there may be more variables at play than simply L1 or L2 status. Assignments in this study were not quality controlled, so the standard of work may have influenced results. Likewise, as alluded to above, cultural and linguistic backgrounds may have had an effect on how the L2S writers used hedges. Overall, what does emerge clearly is the fact that the Published writers in this study generally make common use of a wider range of lexical hedges than the student writers.

In terms of item sharing, levels were generally high, with only one of the 24 sub-corpora (Physiotherapy L2S) showing total sharing of top ten items lower than 80%. Sharing was lowest among L2S writers in seven disciplines. This suggests that these writers make common use of hedges less frequently used by L1S or Published writers. Given that this phenomenon seems to transcend the disciplines, the explanation may be a generally lower level of familiarity among L2S writers with the conventions of lexical hedge use in academic discourse. Hinkel (2005) argues that L2 learners studying in English-speaking countries receive a great deal of exposure to varieties of conversational or informal language, and may tend to employ such features in their writing, as opposed to utilising features valued in academic register. Taking a specific example from this dataset, the hedge *claim* was commonly used by L2S writers in six disciplines, but did not appear in the top ten lists of L1S or Published writers in any discipline. An example from the L2S corpus would be:

'Although, a study conducted in 2000 by Dörnyei and Kormos claims to find no such relation...' (Tesmash2).

The Collins English dictionary (online) defines *claim* in the following way: 'If you say that someone claims that something is true, you mean they say that it is true but you are not sure whether or not they are telling the truth'. Similarly, the New Oxford Dictionary of English states that a *claim* is an assertion made 'typically without evidence or proof'. From an L1 perspective, *claim* can therefore include a connotation of potential dishonesty, or of propositions being made somewhat baselessly, and clearly differs from other reporting verbs such as *argue* and *contend*, which serve to hedge a proposition, but do not question the integrity or honesty of the individual proposing it. It is possible that for some L2 student writers, subtle distinctions

such as this may be less clear - something that could lead to usage that differs from that of L1 student, or expert writers.

It is interesting to note that when overall sharing is considered, in five disciplines the student groups both bore more similarity to the Published group than to each other. This would seem to indicate that while both are making frequent use of hedges common to Published writers, the selection of hedges in question is not the same for both groups of student writers. Why this might be is unclear, although as mentioned above, a number of variables may influence how different student writers make use of hedging devices.

5.4.3 Pedagogical Implications

Hedging, as is widely acknowledged (Hinkel, 2009; Hyland, 2008b; Kim & Lim, 2015; Vasquez and Giner, 2008; Vold, 2006) plays an important role in academic writing, and its value in EAP pedagogy is clear (Hinkel, 2005), but what can the results tell us about the optimal approach to instruction in a context such as a pre-sessional course, where learners may go on to study in a wide range of academic fields? Overall, the choices of lexical hedge use made by the writers in this corpus showed considerable uniformity in terms of the most commonly used items and how they were shared, both across and within disciplines. Nonetheless, important differences were found. Published writers showed the lowest proportion of total hedge use made up by the top ten items in all but one of the disciplines, while L2S writers relied more heavily on modals than did their L1S counterparts. These results may point towards an instructional benefit for L2 student writers if they can be encouraged to depend less frequently on the most familiar lexical hedges. Rates of hedge use among L2S writers were also lowest in six of the eight disciplines. Direct instruction on hedging has been shown to increase both the variety and rates of hedge use by L2 students (Petchkij, 2019), so how is hedging dealt with in

the pre-sessional course? While there were varied recollections, most interviewees remembered having studied hedging on their pre-sessional courses:

 'Yes, because back in Colombia and here in the pre-sessional, every time you say something it's like what is the source of this, or is this true or is this false, or why are you assuming something or can you prove it.' (P1 – five week course)

 'Yes I think so. We cannot say something with *must* or *shouldn't*, we should use *may* or *could*.' (P2 – 10 week course)

 'Oh, I don't think so, I don't really remember, but one of my friends, he studied hedging in an academic English course after the pre-sessional.' (P3 – five week course)

 'Yes of course…they gave us some words and vocabulary, and which is wrong and have compare two words and like this one is good and this one is bad…If I remember correctly, they gave us an article, and we had to find which is wrong and correct it.' (P5 – five week course)

 'Yes, they taught it a lot and they made it very highlighted for academic reasons. They said we cannot just state things without proof, we have always to leave half the door open…For instance I had to write an assignment, based on something that was going to happen, so I always had to hedge - *it seems that, probably, it may be, it may occur that*…and I learned that in the pre-sessional' (P4 – 20 week course)

Nonetheless, while the interview participants all agreed that hedging is an important part of writing in their fields, they felt varying degrees of confidence when asked if they thought they had a good understanding of different ways to hedge in English:

'I would say I can do it basically.' (P1)

'I'm not sure, but I can sometimes. I can use *seems like*, or some different styles of sentences, not just *could, may* and so on.' (P2)

'I have a few ways to use uncertain words, but I'm not really confident in how.' (P3)

'…when it comes to writing it's much easier. Talking is more difficult.' (P4)

'No.' (P6)

Within the existing materials, the Oxford EAP textbook contains a unit which broadly focuses on hedging language. As a concept, hedging is introduced specifically in speaking, rather than writing instruction, but nonetheless the materials demonstrate the use of modal verbs, lexical verbs, adverbs, and prepositional phrases as possible ways to hedge. A later section of the unit, focused on vocabulary, highlights the meanings of lexical verbs including *allege, argue, claim, perceive, suggest*, and *doubt*. Outside of the textbook, hedging and boosting is also covered in the in-house materials, and a variety of modals, lexical verbs (for both speculation and opinion), and adverbs and adjectives (with varying degrees of certainty/uncertainty) are highlighted, including the majority of the lexical hedges appearing most commonly in the various sub-corpora here. Practice activities include gap fills, sentence reformulation, and identifying hedging language from transcripts.

Given that Published writers in this study appear to make use of a wider variety of lexical hedges than the student writers, the existing materials offer an appropriate range of possibilities in terms of demonstrating how different word forms can be used to hedge. However, since L2S writers seem to rely more heavily on modals, one suggestion may be to present a wider variety of lexical verbs for hedging as alternatives. Hyland & Milton (1997) raise a similar point, arguing that there is an over reliance on modals in many textbooks. An

increased focus on lexical verbs would also allow an opportunity to explore meaning more closely, to highlight the sometimes subtle differences that may lead to inappropriate usage, and to draw attention to the differences between formal and informal register, which may also be useful for L2 students (Hinkel, 2005). Given that there is a clear crossover between lexical verb hedging and reporting verbs for citation, there is the added advantage that benefit could be gained in both areas. Further emphasis on the importance of varying the word forms used so as not to depend on modals too commonly may also prove beneficial. This combined approach would arm learners with a variety of lexical options in each word form category, and help to provide a clear focus on how meaning and usage may differ – outcomes that have been recommended (Hyland & Milton, 1997). Learners could be encouraged to practise and produce these varied forms in activities such as paraphrasing texts with differing levels of certainty, free writing, or writing texts loosely grounded in their own subject areas, as opportunities to employ and experiment with the forms they learn are helpful for students (Hyland, 2004).

It is unclear whether the generally lower rates of hedging seen in the L2S writer group represents an issue to be addressed in pre-sessional instruction, although as noted in Section 5.4.2, there is some previous evidence to suggest that L2 writers may underuse hedges when compared to L1 students and expert writers. It is advisable for pre-sessional instructors to ensure that they emphasise the importance of hedging, particularly as their learners may come from cultural backgrounds in which expressing the uncertainty of propositions is less of a convention than it is in academic English (Hinkel, 2005). As Loi & Lim (2019) point out however, this must be done with care if we are to avoid a scenario in which students subsequently over-hedge and qualify unnecessarily – 'a full understanding of hedging devices includes knowledge about the dangers of their misuse' (Varttala, 1999:194). Additional guidance in terms of the

kinds of statements that may or may not require hedging would therefore also be beneficial, and could be given using example texts from a variety of academic disciplines.

Rates of hedging certainly appeared to differ based on discipline, as has been reported previously (see preceding section for details of studies). Again, it must be remembered that this corpus did not include any disciplines from the hard sciences, and so there may be a greater disciplinary variation in hedge rates than is indicated here. Vasquez & Giner (2008) for example, suggest that disciplines driven strongly by precise data, such as mechanical engineering, may require less hedging than those in which results depend more on contextual factors and interpretation, such as marketing. There is no evident attention given in the existing pre-sessional materials to how discipline may influence the need for hedging, and while it may be impractical in a mixed-discipline context to focus too specifically on individual disciplines (Gilquin et al., 2007), it would be possible to present examples from different fields, perhaps, as alluded to above, with very contrasting forms of data. These could be used to focus learners' attention on how the need to hedge may be affected by the kinds of data or the subject being considered, and may, albeit indirectly, help to guide them in reference to their own academic disciplines.

There is no clear indication from the results of this study that the specific lexical hedges used by the writers differ markedly from one discipline to another, with high levels of item sharing in all three writer groups. Indeed the published writers show the greatest commonality in usage across disciplines. From the academic fields in this corpus therefore, differentiating teaching by discipline when it comes to the specific lexical terms for hedging would not be of great value. With that said, the absence of hard science disciplines must again be borne in mind. Vold (2006) observes varied use of epistemic modality markers in linguistics and

medicine research articles for example, and contends that teaching of hedging should reflect disciplines so that students can learn the language of their field. Future analysis of more disciplinarily contrasting academic corpora would be of value to further inform pedagogy in this regard, but as suggested above there is no reason why, in the meantime, example texts from various contrasting disciplines cannot be used as part of pre-sessional materials, with the aim of drawing learners' attention to the ways in which hedging strategies may vary in different academic fields.

5.5 Citation

5.5.1 Variation across Disciplines

The citation forms most commonly used in the corpus were non-integral citations, which outnumbered integral citations by almost two to one. Where integral citations were used, most involved reporting verbs, and discourse act verbs made up the largest proportion of these. These results broadly mirror those of previous studies, where a preference for non-integral forms has been shown (this is particularly true in sciences), and 'soft' disciplines (as opposed to hard sciences) tend to favour discourse act verbs (Hyland, 1999b; Lee et al., 2018; Mansourizadeh & Ahmad, 2012; Okamura, 2008; Samraj, 2013). Looking more closely at the disciplines, an interesting result is that Forensic Psychology was the only discipline where research act verbs predominated among all writer groups, and in Physiotherapy, Published writers made more use of such verbs compared to those associated with discourse. It could be argued that these are the only two disciplines in this study that even begin to approach the 'sciences' end of the disciplinary spectrum. As such, evidence of some preference for research act verbs may not be unexpected, given that this has been noted in engineering and sciences in the past (Hyland, 1999b).

The rate of citation also showed some disciplinary trends. Among all writer groups, the Advertising and Languages disciplines showed the lowest citations per 1000 words. Why this might be so is unclear, although in the case of student writers in Languages one possible explanation may lie in the assignment types. Both of the assignments that formed the sub-corpus for this discipline involved reporting on teaching placements that students had been involved in, and thus sizeable parts of the texts were concerned with relating personal experiences – something that would be unlikely to involve citation. Both student groups also showed their highest rates of citation in Social Work, followed by Forensic Psychology. Indeed, across the disciplines in general there was considerable variation in citation rates among all three groups of writers, although rates varied most in the L2S group, and least in the Published group. While no previous studies involve the same specific range of disciplines as used here, rates of citation have been shown to be affected by discipline. Previous studies have found, among other things, that citation rates tend to be higher in 'soft' disciplines; that Philosophy students used almost four times more citations than students of economics; and that while rates of citation have increased since 1965, this change has not been uniform across disciplines (Ädel & Garretson, 2006; Bahadofar & Gholami, 2019; Hyland, 1999b; Hyland & Jiang, 2019). However, a lack of disciplinary effect on citation density has also been observed in studies such as Hu & Wang (2014), although this study only involved two disciplines.

Considering how the most commonly used reporting verbs were shared across the various disciplines, *suggest* was the most widely shared item, being common to all disciplines in all writer groups. This may be due to its utility not only for reporting a previous writer, but also for serving to hedge what is being reported. Indeed, among L2S writers, all three items occurring in every discipline (also *argue* and *according to)* fit this description. For L1S and

Published writers, the most commonly shared reporting verbs across disciplines showed a mixture of discourse and research focus (L1S – *suggest* and *find*; Published– *suggest, find, show)*. The commonality of such research act verbs among Published writers across disciplines may be a result of the practical research focus of research articles, compared to the varied text types of student assignments.

Sharing overall was generally high in the majority of disciplines for each writer group. Published writers showed a minimum of around 85% in Advertising; L1S a minimum of around 89% in Physiotherapy, and L2S a minimum of around 61.5%, also in Physiotherapy. However, among student writers in particular, many shared items were only common to between 2-4 disciplines, while wider sharing (in terms of number of items) was generally more common among Published writers. Thus, when it comes to citation and RQ1, these results suggest that between the disciplines, there is not a huge amount of variation in terms of which reporting verbs are most commonly used, particularly among Published writers. This group showed the highest total sharing in four of the eight disciplines, although student writers also showed generally high levels of commonality across disciplines. Among Published writers, a maximum of two items from the top ten remained unshared in any discipline (Advertising). This was slightly higher for L1S writers (three of 14 items in Fashion) and higher again for L2S writers (9 of 16 items in Fashion)[6] It may be that the higher shared usage and fewer unshared items seen among Published writers result from a greater familiarity with academic writing and the reporting verbs commonly used in this genre. While these results suggest that disciplines to not differ notably in terms of which reporting verbs are used most commonly in research

---

[6] Top ten lists were compiled on the basis of proportion of total integral citation use, so may exceed ten items if a number of verbs appear in the same quantity.

articles, it is possible that more variation across disciplines would be found if hard science subjects were included in the analysis. Hyland (1999b:349) notes that there is considerable disciplinary variation in terms of verb use, and goes as far as to argue that his data suggests 'writers in different fields almost draw on completely different sets of items to refer to their literature'. This was not something seen in this study, and may be due to the absence of science disciplines. It seems overall, that student writers show somewhat less uniformity across disciplines than the Published writers, but whether this is a result of a genuine lack of disciplinary variation in research articles is hard to say.

5.5.2 Variation within Disciplines

Importantly, Published writers showed the lowest rates of citation in the majority of the disciplines. While this would seem to contradict the findings of previous work such as Kafes (2017) and Mansourizadeh & Ahmad (2011), which found higher densities of citation among expert writers than among novices, it is important to note that these studies both centred entirely around research articles, and as such their results cannot be extrapolated to student assignments. One possible explanation for the higher citation rates among student writers may be a tendency to over-cite. Student writers may do this in an effort to avoid plagiarism accusations, or to demonstrate a wide range of reading so as to gain additional marks (Davis, 2013; Harwood & Petrić, 2012; Petrić, 2012). It may be the case that in assignment types not involving primary research, students must rely solely on previous findings, and thus require a greater density of citations. They may also feel to a greater extent than the published writers, that when they *do* make their own assertions, these need to be validated by referencing previous work, given students' relatively 'novice' academic status. Among the students themselves, while L2S writers showed the highest citation rates in five of the eight disciplines,

the differences in rates between them and their L1S counterparts were sometimes quite minimal, and in the majority of disciplines, less pronounced than the differences observable between students and published writers. Some studies, such as Borg (2000) have argued that L1 and L2 students can encounter similar issues when it comes to using citation effectively. Rather than being an issue of language familiarity therefore, the differences in citation rates here may be influenced to a greater extent by the audience these writers aim their work at, and the position they perceive themselves to occupy in their academic discourse communities. Citation among student writers may be more centred around knowledge display, citing those with a higher academic standing. Conversely, published writers cite their peers, and may do this for a variety of reasons, such as to express 'allegiance to a school of thought' (Petrić, 2007:239) or to show the significance of a study (Mansourizadeh & Ahmad, 2011). Therefore, even if student writers, as Pecorari (2006) suggests, take the publications they read as exemplars of how academic writing should be done, there may be differences in what they feel constitutes appropriate citation when compared to published writers. It is important to remember that assignments and research articles have entirely different audiences and purposes.

 The picture of integral and non-integral citation use within disciplines was a mixed one. In four of the eight disciplines Published writers showed the highest proportional use of integral citations. L2S writers showed the lowest in four disciplines. Meanwhile L1S writers showed the highest integral use in one discipline, and the lowest in two[7]. Integral citations place more emphasis on the author being cited, while non-integral citations give emphasis to the findings

---

[7] Published writers used over 95% non-integral citations in Physiotherapy, but this was due to the superscript referencing system in these RAs.

themselves (Hyland, 1999b). Thus, using one or the other constitutes (at least in theory) a rhetorical choice. However, many factors have been found to influence the choice of citation form, including discipline, genre, language, and writer experience (Jomaa, 2019), and there is some evidence to suggest that novice writers may lack awareness of the rhetorical functions of different citation forms (Jalilifar, 2012a; Nguyen & Pramoolsook, 2015; Okamura, 2008). Among student writers, and L2 student writers in particular in this study, the proportionally greater use of non-integral citations when compared to published writers may, like citation rates in general, be influenced by the position they occupy in their discourse communities. Hewings et al. (2010) contend that non-integral citations allow a writer to acknowledge others' work and thereby demonstrate that he or she is knowledgeable, in the process improving his or her academic credibility. This, it is not hard to imagine, would be among the primary concerns for postgraduate student writers, seeking to demonstrate understanding of their field and an appreciation of the previous work within it. Integral citations, in foregrounding previous authors, allow writers to position themselves within their field, for example by aligning themselves with previous arguments or contentions, which may be more of a concern to published writers. This may go some way to explaining the generally higher use of this form among the Published writers in a number of the disciplines. Excluding Physiotherapy, only one discipline (Forensic Psychology) showed Published writers exhibiting the highest proportion of non-integral citation use while L2S writers exhibited the lowest. Why this discipline should differ is unclear, although as described in Section 5.5.1, it does show other citation patterns more commonly associated with science-oriented subjects, so perhaps in this too, there is a disciplinary influence.

Examining the reporting verbs and how their usage differed within the disciplines, it is clear that Published writers make common use of a wider variety of reporting verbs than either of the student groups, given the lower proportion of total integral citation use accounted for by the top ten verbs among this group. Again however, there is a mixed picture when it comes to the student writers, with L2S showing the highest proportion in only five of the eight disciplines. Here too, it appears that L1 student writers, rather than overcoming the difficulties of English academic writing simply by virtue of their language background, do face some of the same issues encountered by their L2 counterparts. Having said that, the L2S writers in all but one discipline showed notably higher use of *according to* in their integral citations than did either of the other writer groups. This may be because, as a syntactic structure, this is relatively straightforward to use, and lexically does not require the selection of an appropriate reporting verb such as *contend, argue*, or s*uggest*. L2 writers in particular may therefore find this a simple and reliable means of structuring an integral citation. However, the fact that in three of the disciplines *according to* constituted close to or over 30% of all integral citations for L2S writers, would suggest some over-reliance on this structure.

No clear patterns emerged in terms of how writer groups shared reporting verbs, and disciplines varied widely in terms of which writer group shared the largest or smallest percentage of their top ten usage with other groups. What did emerge was that for L2S writers in four, and L1S writers in five disciplines, student writers bore more similarity to each other in terms of reporting verb usage than they did to the Published writers. Combined with the findings regarding citation density, integral vs. non-integral forms, and reporting verb variety, this further suggests that the predominant differences in citation practices in this study arise not from L1 or L2 status, but from whether a writer is a student or a published academic.

5.5.3 Pedagogical Implications

Appropriate citation in academic writing is important. It helps to provide context, outline the problem a piece of work contributes to, addresses, or arises from, and assists readers in evaluating the existing knowledge within a specific discipline (Charles, 2006; Hyland, 1999b). For student writers, citation also serves to demonstrate to lecturers the individual's subject knowledge and an understanding of what has come before in their academic field. It is therefore important that all students have a good understanding of citation practices, and EAP contexts such as pre-sessional courses can play a vital role in this where L2 students are concerned. The results give some interesting indications in terms of which aspects of this complex practice might best be focused on.

The interviewees noted varying levels of confidence in their ability to utilise citations correctly and effectively, and highlighted a variety of difficulties they encounter in attempting to do so:

'I do feel confident doing it, because I used to do it in Colombia…I knew the rules before, and they covered it in the pre-sessional, and as far as I know they covered everything you need to know.' (P1)

'No, I don't think so. I have confidence to write regular references…I know how to write the structure, but I don't know what I can use and what I can't use.' (P2)

'I might make some mistakes, but I'm confident, because we have been taught about it on the pre-sessional' (P4)

'No…because in the pre-sessional class, we changed the tutor, so I still remember one of the teachers said the reference should be written in this way, and another one said, "No, this is wrong", so until now I'm still confused about how to write a reference in the correct way.' (P5)

'I remember that the pre-sessional took more than two classes to talk about this, but I think it's very hard for students to remember' (P6).

Examining the existing pre-sessional materials confirms that a great deal of attention is given to citation practices. The Oxford EAP textbook *and* the in-house materials include a specific focus on integral versus non-integral citations. Both formatting and the notion that these two styles serve to create different emphasis (either on the content cited, or on the author) are explained; however, the rhetorical purposes of these two citation forms appear only as 'information boxes', and this idea is not a focus of subsequent practice. A variety of reporting verbs are presented, including *argue, believe, say, state, suggest, report, describe, claim, assert, stress and maintain,* and there is guidance in terms of which of these might be most appropriate in circumstances such as making an argument, agreeing/disagreeing with a point of view, and commenting on evidence or statistics. More widely, the materials also give attention to different referencing strategies, such as summarizing, paraphrasing, and direct quotation, and discuss when each of these may be most appropriate. Students are given the opportunity to practise these various aspects of citation use through activities such as gap fills, identifying and analysing citations from example texts, reformulating sentences (integral to non-integral and vice-versa, for example), and writing referenced paragraphs.

While, as outlined above, many of the important elements of citation are well-covered in the pre-sessional materials, there may be issues in terms of how this knowledge and information is carried over from the pre-sessional course to the students' subsequent studies. Some students may simply revert to their previous study habits, and it has been argued that EAP should be 'directly involved with the pedagogy of source use on subject programmes, so that students achieve this transfer and continue developing the skills they need' (Davis, 2013:134).

Taking integral and non-integral citation use as an example, despite covering the rhetorical differences between these forms in their pre-sessional courses, some interviewees were clear that this is not what now guides their choice of which citation form to use. P1 commented that for him, varying how he writes citations is simply a matter of not wanting to use the same format all the time. P2 acknowledged that she used the same form most of the time, but was confused as she had been told different things by different lecturers - even being told that citations must be non-integral and come at the end of sentences - something in direct contradiction to pre-sessional content. Similarly, P5 and P6 said that they usually write citations the same way. Only P3 mentioned any rhetorical purpose in her choice of integral or non-integral citation, explaining:

 'If I talk about background information, I'm going to use like "blah blah blah with (Smith, 2007)", but if I am talking about some theory the author mentioned, I'm going with "Smith (2007) blah blah blah."'

 Accepting the notion, elaborated upon in Section 5.5.2, that many factors may affect the choice of integral or non-integral citation, the Published writers in this study did show higher proportions of integral citations than the L2S writers in five of the eight disciplines. This, combined with the interviewees' comments above, might be an indication that pedagogical benefit could be drawn from focusing increased emphasis on the rhetorical purposes of citation forms, particularly as the majority of practice activities in the existing materials serve predominantly to address format rather than the reasons *why* a given form might be chosen. This could perhaps be achieved through means of activities such as matching citation forms with their rhetorical function, or analysing citations in texts from the point of view of the writer's rhetorical intentions (Petrić, 2007). With a greater emphasis on the rhetorical aspect,

students may be more likely to carry these ideas across to their courses of study, and use them in their academic writing.

In terms of citations rates, there is some indication of disciplinary variation consistent between writer groups (Advertising and Languages having the lowest rates for all writers), although the variation in rates across disciplines was least pronounced for published writers. Within disciplines the student writers (L2S in particular) tended to display higher rates of citation than the published writers. While this could, as discussed in Section 5.5.2, simply be the result of students and RA authors writing for different purposes and audiences, and therefore not represent a teaching need, there may be some value in the pre-sessional course addressing, in addition to the importance of citation to avoid accusations of plagiarism, the notion of over-citing. It is not difficult to see how warnings such as this from the in-house materials may encourage students to err on the side of citing as much as possible, rather than employing a more nuanced view of when citation may and may not be necessary:

> '**Warning!** It is important to cite sources correctly. If you do not, you may be committing plagiarism. Plagiarism is the practice of taking another person's work or ideas and presenting them as your own. This is considered academic misconduct and can result in expulsion from university.'

(p. 90)

It would perhaps be possible to integrate this into existing sections of the materials that deal with the appropriacy of different kinds of source use, such as paraphrasing, summarising, and direct quotation. The idea that citation may play a more important role in some disciplines than in others may also be a pertinent one. Hyland (1999b:353) notes differences in citation rates between 'hard' and 'soft' disciplines, suggesting that the framework of theoretical knowledge in hard sciences results in relatively fewer citations compared to social sciences and humanities, in which 'issues are more diverse and detached from immediately prior

developments…Writers draw on a literature that often exhibits greater historical and topical dispersion'. This, he suggests, results in the need for more frequent citation. Nonetheless, while it is important for students to have an appreciation of the conventions of their own disciplines, pre-sessional courses may not always be the ideal context, and this may be one area in which instruction on individual academic course programmes would be more appropriate, particularly given the subject familiarity necessary to advise students on how citation may be approached in specific disciplinary discourse communities.

When it comes to the variety of reporting verbs used in integral citations, and how these may be affected by discipline, rates of item sharing across disciplines were generally high (above 80% in all but two of the 24 sub-corpora). Again however, as with the other academic writing features examined in this study, it must be remembered that the corpus did not include any hard sciences, and this may have restricted the extent to which any disciplinary variation in citation use was noticeable. A number of previous studies have pointed to the importance of citation instruction that accounts for disciplinary differences (Charles, 2006; Ho & Hang, 2014; Mansourizadeh & Ahmad, 2011; Petrić, 2007; Shi, 2008), but the present results do not reflect such a need, and further studies would be necessary to clarify this.

Published writers made use of a wider variety of reporting verbs than the student writers, and although the existing materials do include, as mentioned, a variety of possible verbs for citation, there may be scope for a broader focus here. Hyland's (1999b, 2002b) taxonomy of reporting verbs as research, cognition, or discourse related could represent a useful starting point. This kind of differentiation is not mentioned in the existing materials, but if learners were presented with a variety of examples of reporting verbs that fit these different categories, this might encourage them to make more diverse use of such verbs, rather than relying so

heavily on perennial favourites such as *state, suggest, find,* and *argue*. It would also help to raise their awareness of these different functions, allow for the possibility of reporting verb use more closely tied to the particular content, arguments, or procedures being cited, and provide, as Hu & Wang (2014) suggest, resources to construe various writer stances - something that would benefit students not only in citation but also in the closely related practice of hedging. Such instruction, and a greater emphasis on the importance of employing a variety of reporting verbs in writing may also help to address issues such as the over-reliance on *according to* demonstrated by some of the L2S writers in this study.

6 Conclusion

6.1 Overview of Chapter

This chapter summarises and brings together the main points of the discussion in order to present the overarching conclusions of the study in terms of the pedagogical implications for the specific pre-sessional context, and for EAP instruction more widely. Additionally, the possible limitations of the study are considered, and recommendations for future research in this important area are proposed.

The study aimed to help inform academic writing pedagogy in the specific context of a UK university pre-sessional EAP course, by examining four important features of academic writing – vocabulary, lexical bundles, hedging, and citation. This was achieved by building and analysing an academic corpus consisting of assignments by Level 7 L1 and L2 students and research articles by published authors, across eight academic disciplines. The four features were analysed both across and within disciplines in order to highlight similarities and contrasts between both the academic disciplines and the different writer groups, and the corpus analysis was combined with insights from interviews with past pre-sessional participants, and an examination of the course's existing materials. These analyses served to address the four research questions, and the findings were then applied to the specific context of the institution's pre-sessional EAP course.

RQ1. What is the extent of disciplinary variation in the use of academic vocabulary, lexical bundles, hedging and citations?

The extent of disciplinary variation seen in the four features was not uniform. Citation and hedging generally showed notably lower levels of variation across the disciplines, compared to either lexical bundles of academic vocabulary.

RQ2. How do L2 student, L1 student, and published academic writing differ in the usage of academic vocabulary, lexical bundles, hedging and citations?

In both vocabulary and lexical bundle usage, the L1S and L2S writers bore more similarity to each other than to the Published writers when it came to the most commonly used items, while for hedging and citation use the picture was less clear. However, an important observation is that, in both hedging and citations, published writers relied less heavily on the most commonly used items than did either of the student groups.

RQ3. How are academic vocabulary, lexical bundles, hedging and citations represented in existing pre-sessional writing materials?

The most notable finding in the analysis of the existing pre-sessional materials is the lack of attention given to lexical bundles, which are not explicitly mentioned and do not form the basis for any explicit teaching. The other features are all represented in the existing materials, but there is potential for adaptations and improvements to be made.

RQ4. What are the perceptions of students who have completed the pre-sessional course in terms of the teaching of these four academic writing features, and their own difficulties with academic writing?

The absence of lexical bundles and subject-specific academic vocabulary were both noted by the students. While the majority of interviewees recalled being taught both hedging and citation use during the course, the extent to which they felt this had prepared them for their

subsequent studies varied. Difficulties with academic writing highlighted by the students included recognising academic vocabulary and using a sufficiently varied academic lexicon, hedging effectively, having confidence in their ability to cite correctly, and knowing the rhetorical significance of different citation forms.

6.2 Overall Conclusions for EAP Writing Pedagogy

6.2.1 Vocabulary, Lexical Bundles, Hedging, and Citation

EAP courses generally aim to improve academic language proficiency and teach the necessary study skills, given that L2 students can face particular difficulties adapting to writing academically in their disciplines, be this mastering subject-specific vocabulary, awareness of register, use of cautious or hedged language, effectively acknowledging sources, appropriate writer voice, or meeting the demands of their particular audience (Chen & Baker, 2010; Dong, 1998; Gilquin & Paquot, 2007; Ivanič & Camps, 2001; Jordan, 2002, Terraschke & Wahid, 2011). Aspects of all four writing features were identified as problematic by the interviewees in the study, and through their insights, combined with findings from the corpus and an examination of the current materials for teaching writing, it has been possible to make a number of recommendations with particular reference to a specific EAP pre-sessional course. However, these are also of relevance and likely benefit to the wider HEI pre-sessional context, as L2 students at other institutions will face similar problems with English academic writing, and must meet similar demands in terms of what is expected of them once they begin their studies.

Overall, three of the features – academic vocabulary, hedging, and citation – are well-represented in the existing pre-sessional writing materials, although in all three cases, the results have led to additional recommendations. For vocabulary, the importance of encouraging the use of a wider variety of academic vocabulary is emphasised. Given the

practical constraints of the pre-sessional context, and the potential advantages for the students themselves going forward in their academic lives, this may best be achieved through instruction in vocabulary learning strategies - raising learners' awareness of strategy use, presenting them with a variety of possible strategies, and allowing ample opportunity to use them, evaluate them, and adapt them. This could be combined with the use of disciplinarily differentiated materials, such that learners can not only be exposed to some of the subject-specific vocabulary of their own fields, but can be encouraged, through consciousness-raising tasks and strategy use, to further explore such vocabulary independently. This would go some way to acknowledging the importance of teaching for discipline in EAP vocabulary, as argued by many researchers (Durrant, 2014; Hyland & Tse, 2007, 2009; Vongpumivitch et al., 2009, for example), but would not require that large amounts of class time be devoted to the teaching of such vocabulary in a context where English for general academic purposes is often more practicable than English for specific academic purposes (Gilquin et al., 2007). Focusing on the acquisition and production of academic vocabulary through improving learner strategy use and autonomy would also have the advantage of not placing the burden on EAP instructors, who may not have the subject knowledge necessary to effectively teach discipline-specific vocabulary in any case.

For hedging, it is recommended that materials highlight a wider range of lexical verbs (which would also feed into citation instruction, as many of the lexical verbs for hedging also serve as reporting verbs for citation) and also word forms, again with the aim of encouraging learners to diversify their usage and to rely less heavily on modal verbs. Additionally, it may be beneficial, while raising awareness of the importance of hedging, to introduce a nuanced view, if learners are not to over-hedge. This could include highlighting disciplinary differences in

terms of how necessary it may be to hedge given different kinds of data, for example hard,

quantitative data versus more context-dependent qualitative data. Disciplinarily differentiated

tasks and materials could again be used here to contrast such differences in how varying

academic fields may employ hedging, without sacrificing the general applicability of the

content. As Hyland (2002a:393) observes:

> A major problem of heterogeneous classes is actually finding enough common ground among students, but one solution is to exploit the specificity of their circumstances through the opportunities that such classes offer to contrast their disciplinary experiences and expectations…this kind of rhetorical consciousness raising not only helps satisfy students' demands for personal relevance, but also reveals to them the multi-literate nature of the academy.

Citation in the existing materials receives detailed and thorough attention. Nevertheless,

benefit could be derived from an increased focus on the rhetorical functions of integral and

non-integral citations, to complement the focus on form and structure. This could include tasks

based on ideas such as reformulating citations to fulfil a specific rhetorical purpose, or

identifying the rhetorical purpose of example citations. Such awareness raising may encourage

learners, when they make a choice about which citation form to employ, to do so based more

on the rhetorical purpose of a given form, than on simply varying it for the sake of varying it,

or even using the same form the majority of the time. If learners can be more aware of the

different functions of citations, they can apply this knowledge to their linguistic choices

(Charles, 2006). Additionally, more nuanced and varied use of reporting verbs for citation could

be encouraged by placing more emphasis on and raising awareness of the different categories

of reporting verb – discourse, cognition, and research oriented. Presenting examples and

allowing learners to explore, revisit, and use them productively may help them not only to

diversify the forms they utilise, but also to consider more deeply what kind of information they

are citing (for example, statistically significant quantitative findings, qualitative interview data, argument or opinion, research procedures and so on) and what the most appropriate kind of reporting verb might be. As mentioned above, an increased awareness of the potential variety of reporting verbs for citation would also link beneficially to hedging, since there is substantial crossover between these two areas.

The only feature which does not already receive attention in the existing pre-sessional writing materials, is lexical bundles. Various studies have argued that bundles represent an important element of EAP pedagogy, and that identifying and prioritising those that may prove most useful to learners is important (Bychkovska & Lee, 2017; Chen & Baker, 2010; Hyland, 2008a, 2008b; Simspon-Vlach & Ellis, 2010). It is recommended that lexical bundles therefore be included as a focus in pre-sessional writing instruction. Tasks and materials that serve to raise awareness of bundles would be beneficial, particularly consciousness-raising tasks that facilitate the retrieval, use, and manipulation of items, and help learners to employ them in extended writing (Hyland & Tse, 2009). Authentic texts from a number of academic disciplines could be used to highlight and contrast those core bundles that may be useful in a wide variety of fields, with those that may be more specialised, thus again serving the needs of the learners as a heterogenous group, while also acknowledging disciplinary variation. In the longer term, further research on academic corpora from a variety of disciplines, produced by both students and expert writers would greatly help to identify the most important lexical bundles for EAP pedagogy. Of course, frequency is only one factor in determining what might be included, but nonetheless, CL can be an important element in the 'selection, sequencing and structuring of teaching content' (Hyland, 2008b:60). In the shorter term, existing materials could be meaningfully adapted with reference to the existing literature, and studies such as this one.

6.2.2 On Disciplinary Variation

While there is certainly an argument, particularly in the literature, for firm disciplinary specificity in EAP, as alluded to previously, this may not always be a realistic or practicable approach in pre-sessional contexts. The results here also suggest (accepting that the overall picture of variation across disciplines is limited by the absence of hard sciences in this study) that disciplinary variation may well be more important in some areas than others – lexical hedges seem to show far more commonality between disciplines than do lexical bundles, for example, and variation in a number of instances, was greater among student than published writers. If teaching that is differentiated for discipline is not required for all areas of academic writing, or even for all aspects of a given feature of academic writing, then to what extent should pre-sessional courses employ materials and teaching specific to disciplines, particularly when disciplinary boundaries can, in any case, be porous (Durrant, 2017)?

Hu (2007) acknowledges the potential dilemma for course designers - even though students are expected to be able to meet the demands of their academic courses, and therefore writing instruction should ideally be specific to those disciplines, practical constraints often make this impossible. He goes on to suggest however, that despite disciplinary variation in academic writing, there are reasons for courses such as pre-sessionals to remain predominantly focused on English for general academic purposes. Firstly, he contends that the tasks required of students in different disciplines share many characteristics (although Hyland (2002a) cites a number of studies that would call this notion into question). Secondly, effective writing skills, such as information gathering, planning, editing and revising, and other strategies, can be applied across university curricula. I would argue that, as reflected in the recommendations made in the preceding section, a pragmatic approach to acknowledging disciplinary variation

– one that neither discounts it as impractical, nor demands that it form the basis of EAP writing pedagogy – may be the most realistic and helpful. It is true that some UK HEIs do offer discipline-specific pre-sessional courses, such as English for business, engineering, or education, made in collaboration with the relevant university faculties (University of Nottingham, 2020), but pre-sessional participant numbers may not make this feasible in all institutions. Moreover, while one of the interviewees in this study commented that she would like to see pre-sessionals split in this way, a number of the others, while acknowledging that such courses might be useful in preparing them for their disciplines, expressed a preference for pre-sessionals that are mixed-discipline, saying that they valued being able to learn and communicate with students of other academic disciplines. It may be that much can be achieved without the necessity of dividing pre-sessional learners by discipline, particularly if they can be given exposure to some subject-specific materials, which enable them to appreciate both the commonalities and contrasts between disciplines; can be offered opportunities to write on subjects related to their field; and can have their own strategy awareness and use encouraged and raised.

There are of course other possibilities that institutions may wish to explore in the future. These might include ideas such as combining in-class, general EAP, with more discipline-specific online course components, or, as Wingate & Tribble (2011) suggest, changing the model such that academic writing instruction becomes embedded in the curricula of individual disciplines, with subject lecturers playing a more prominent role in teaching it. However, in the meantime, it is entirely possible to make a number of relatively small adjustments to materials and pedagogy in order to more fully address the academic writing features explored in this study, take account of disciplinary variation, and potentially benefit L2 university EAP learners.

6.3 Limitations of the Study

6.3.1 The Corpus

While the corpus as a whole was extensive in terms of overall word count, there were some issues within the various sub-corpora. A number of these - those of the Physiotherapy discipline in particular - were relatively limited in size, and thus generated either no results (in the case of identifiable bundles in the Physiotherapy L2S sub-corpus for example), or results that were limited in the extent to which generalisations could legitimately be drawn from them, given the small pool of assignments/writers that constituted the sub-corpus. This imbalance in the size of the sub-corpora was regrettably unavoidable given the realities of building the corpus; the number of students from whom consent could be obtained was, even with best efforts, beyond my control. However, if particularly small sub-corpora could have been avoided, this would have been beneficial.

The corpus in this study contained eight disciplines, which was an adequate number in order to investigate variation, although the original intention was to include 15 – three from each of the five university faculties. Had this been possible, it would have provided a more balanced picture of disciplinary variation, particularly where the contrast between hard sciences and other disciplines is concerned. Previous studies (Cortes, 2004; Hardy and Römer, 2013; Hyland, 1999a, 2008a; and Vázquez and Giner, 2008 to name a few) have highlighted variation between science/engineering subjects and those within the humanities/business sphere, and it would have been useful to re-examine this variation and its pedagogical implications in the present study. Unfortunately it was not possible to include any courses from the hard sciences, and Physiotherapy was probably the closest any of the disciplines in this study came to being science related. Were the study to be carried out again, a more balanced corpus including a

wider variety of science, engineering, and business subjects would provide a more complete picture of the full scope of variation in university academic writing.

As noted in the discussion, assignment topic may have influenced some of the results in the study. The practicalities of data collection meant that some of the student sub-corpora consisted entirely of responses to a very limited range of topics, and this may have affected the outcomes of the corpus analysis, particularly in terms of comparisons with the published research articles, which were not similarly constrained. Ideally, the student sub-corpora in each discipline would consist of a similar range of topics to that of the published sub-corpora, so as to minimise any possible influence from this factor. However, without corpus building on a significantly larger scale than was possible in this study, controlling for topic in this way is simply unfeasible.

6.3.2 The interviews

As a result of a limited response from potential participants, the interviews conducted for this study were few, and thus provided only limited, if nonetheless useful, insights into the learners' perspective on academic writing and pre-sessional EAP instruction. With only six interviewees, and the majority of these studying the same Level 7 discipline, the picture that could be built was not as complete as had been hoped. Were the study to be repeated, interviews with a larger sample of previous pre-sessional participants would be preferable. Ideally, this would encompass a wide range of Level 7 courses, and a healthy distribution of individuals from each of the pre-sessional courses (5, 10 and 20 weeks). This would allow the qualitative element of the study to be more prominent and to provide a more effective complement to the corpus data.

6.3.3 Data Analysis

At various stages in the process of data analysis it was necessary for me to make judgements on the inclusion/exclusion, and categorisation of data. This included instances such as cases of potential lexical hedging, where a given example was ambiguous in terms of whether it should be classified as a hedge, or was being used in some other sense, and decisions over the classification of reporting verbs as discourse, cognition, or research related. While the vast majority of cases were clear, and I am satisfied that I employed a consistent system for those that were less so, it would still have been advantageous to have access to second opinions where these judgements were concerned, even if only to confirm that the context provided by the concordance lines could not remove the ambiguity. For this reason, suitably experienced and qualified assistants would aid in allowing these judgements to be made with added confidence.

It is also important to acknowledge that while, as detailed in Section 3.8.1, the AWL was used in this study for a variety of reasons, there is cause to question whether this offers a representative picture of a general academic vocabulary, even if one accepts that such generality exists. The list has been shown to offer more coverage in some disciplines than others (Hyland & Tse, 2007), and has also been criticised for its focus on word families, wherein no account is taken of grammatical parts of speech, and the members of which may not share core meaning. Additionally, the AWL's relationship with the GSL has been criticised, since this is an old list, no longer reflecting high-frequency English (Gardner & Davies, 2014). While this study has only sought to compare the various sub-corpora to each other, rather than make any definitive judgements about the extent or nature of academic vocabulary within them, it must nonetheless be made clear that any picture of academic vocabulary from this study is based

on analysis using the AWL, and the limitations of this tool should therefore be acknowledged. It is possible that a different picture of academic vocabulary might emerge if the study were to be repeated using an alternative such as the AVL or NAWL. Indeed, this would be an interesting comparison in itself.

The final analyses of the four features in the study were restricted to the most commonly occurring items. This was done for a number of reasons. Firstly, the items that are used most frequently across writers and disciplines are also those most likely to be relevant for pedagogy. Secondly, with such a large amount of potential data (over 400 different reporting verbs for citation alone, for example), it was felt that restricting the analysis to only the most common items, would optimise the utility *and* manageability of the dataset. Nonetheless, it must be acknowledged that this approach excludes a lot of data, from which further insights into how disciplines and writers may differ could potentially be gained.

6.4 Directions for Future Research

There were many additional aspects of the four features of academic writing examined in this study that simply could not be investigated here due to the constraints of time and word count. Additionally, the study suggests directions in which research may beneficially move in the future.

In terms of the four features of academic writing examined in this study, there are a number of interesting foci for potential future research. Work continues into identifying a 'core' academic vocabulary, but as Malmström et al. (2018) point out, this is a complex task involving many variables. If we are to develop as comprehensive a picture as we can of how academic vocabulary is used in universities, these variables must be considered in any future CL studies aiming to inform pedagogy. For example, one potential route may be to consider differences

not only between disciplines and writers, but also between productive and receptive vocabulary – are the words L2 students need in order to write successfully, the same as those they need in order to read, and if not, how might EAP pedagogy approach this?

While lexical bundles have been the subject of a great deal of research, there are, to my knowledge, few (if any) studies that have combined attempts to identify the most pedagogically relevant bundles for EAP pedagogy, with proposals for or investigations into how such bundles might best be taught in contexts such as pre-sessional courses. Integrating these strands of research may prove extremely useful in optimising pedagogy where this important academic writing feature is concerned. Additionally, as noted in Section 5.3.3, more work on corpora including a wider and more diverse range of disciplines needs to be done in order to help assess disciplinary variation more fully and identify those bundles most useful for EAP purposes, be this 'core' bundles, or those specific to different disciplines. Further research might also consider how discipline affects the kinds of bundles used, in terms of Hyland's (2008b) taxonomy of research, text, or participant-oriented, as this would offer deeper insights into the most pedagogically useful bundles.

This study drew some useful insights from examining lexical hedges, but how student writers employ epistemic modality is a complex issue, and does not depend solely on which academic discipline they are writing in. Studies such as Hinkel (2005, 2009), Steinman (2003), and Uysal (2012) have noted how different languages and cultural backgrounds have varying attitudes and approaches to evidence, expressing doubt, and making claims forcefully. It may be helpful for future studies to consider L1 and background in any examinations of L2 student writing, as the potential for this to inform pedagogy (particularly for hedging, but also perhaps for other aspects of academic writing) may be considerable. This is not to suggest in any way that

teaching be delineated on the basis of learners' cultural or linguistic backgrounds, but a fuller understanding of how these factors may affect aspects of academic writing such as hedging, may allow EAP instructors to raise awareness among all learners of potential issues that may arise.

One element of citation that was not examined in this study, but which could have important implications for pedagogy, having been linked previously with higher scores in MA theses (Petrić, 2007), is the evaluative element of reporting verbs. This refers to Hyland's (1999b, 2002b) categorisation of these verbs as 'factive', 'counter-factive', or 'non-factive'. Studies comparing how L2 and L1 student, and published academic writers employ these evaluative functions would provide further insight into how EAP pedagogy should best approach the teaching of citation.

More generally, if we are to build a truly comprehensive understanding of how discipline affects features of academic writing such as those examined in this study, we need, as Vold (2006) argues, to examine not only a wide variety of disciplines, but also sub-disciplines. While such thorough understanding is lacking, judgements on how EAP pre-sessional courses tackle issues such as disciplinary variation, 'core' vocabulary and lexical bundles, and disciplinarily differentiated materials and teaching, will only be best guesses.

While the focus in terms of pedagogy in this study has been L2 learners and pre-sessional courses, the results have additionally highlighted the notion that L1 students may also encounter problems with academic writing. This would be another useful area for future study. Much of the previous research has concentrated on L2 students and their English academic writing, presumably because of the obvious teaching need and the fact that EAP predominantly serves these learners. However, many L1 students are also novices when it

comes to writing academically, and may encounter many of the same difficulties as their L2 counterparts. Further studies into L1 student writing, the problems these writers perceive themselves to have, and the ways in which their writing may affect their academic success would help to make clear just how much of a teaching need exists here. This would, in turn, give HEIs a clearer insight into the potential benefits of rethinking academic writing instruction, moving away from the notion that this is something predominantly of benefit only to L2 students, and towards a model in which academic writing instruction is also integrated within academic courses, and can thus serve both L2 *and* L1 students. If such a model were found to be worthwhile, this would also help to remove many of the issues surrounding EAP and discipline, since subject-specific elements could be addressed in-context, leaving EAP instructors to focus their efforts with L2 learners on more general aspects of academic English.

This study has made a number of useful recommendations for EAP writing pedagogy in the HEI context, but further research, such as that outlined above, is necessary if we are to build a more fully comprehensive understanding of university academic writing in general, how features such as those examined in this study are used within it, the difficulties faced by L2 (and indeed L1) student writers, and the ways in which EAP courses can seek to address such issues and optimise teaching outcomes. Ultimately, these are complex and interconnected issues, but are also of vital importance for both students and higher education institutions.

References

Abasi, A. R., Akbari, N., & Graves, B. (2006). Discourse appropriation, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school. *Journal of Second Language Writing*, 15(2), 102-117. https://doi.org/10.1016/j.jslw.2006.05.001

Abdolrezapour, P., & Tavakoli, M. (2013). University teachers and students' perceptions of EAP methodologies and their effectiveness. *The Social Sciences*, 8(1), 49-54. http://docsdrive.com/pdfs/medwelljournals/sscience/2013/49-54.pdf

Abdullah, F. S. (2005). Analysing EAP needs for the university. In S. H. Chan & M. E. Vethamani (Eds.), *ELT Concerns in assessment*, (202-233). Sasbadi Sdn. Bhd. https://www.academia.edu/download/7036025/Faiz%20Analyzing%20E(S)AP%20Needs%20for%20the%20University_Book%20chapter%20(2005).pdf

Ädel, A. (2006). *Metadiscourse in L1 and L2 English (Vol 24)*. John Benjamins. https://doi.org/10.1075/scl.24

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. https://doi.org/10.1016/j.esp.2011.08.004

Ädel, A., & Römer, U. (2012). Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics*, 17(1), 3-34. https://doi.org/10.1075/ijcl.17.1.01ade

Ädel, A., & Garretson, G. (2006). Citation practices across the disciplines: The case of proficient student writing. In *Academic and professional communication in the 21st century: Genres,*

rhetoric and the construction of disciplinary knowledge. *Proceedings of the 5th International AELFE Conference* (pp. 271-280).

http://www.unizar.es/aelfe2006/ALEFE06/1.discourse/38..pdf

Alsaawi, A. (2014). A critical review of qualitative interviews. *European Journal of Business and Social Sciences*, 3(4), 149-156. https://doi.org/10.2139/ssrn.2819536

Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71-83. https://doi.org/10.3366/e1749503209000227

Andrade, M. S. (2006). International students in English-speaking universities: Adjustment factors. *Journal of Research in International Education*, 5(2), 131-154. https://doi.org/10.1177/1475240906065589

Ang, L. H., & Tan, K. H. (2018). Specificity in English for Academic Purposes (EAP): A corpus analysis of lexical bundles in academic writing. *3L The South East Asian Journal of English Language Studies* 24(2), 82-94. https://doi.org/10.17576/3l-2018-2402-07

Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Aull, L. (2015). *First-year university writing: A corpus-based study with implications for pedagogy*. Palgrave Macmillan. https://doi.org/10.1057/9781137350466

Aull, L. L., Bandarage, D., & Miller, M. R. (2017). Generality in student and expert epistemic stance: A corpus analysis of first-year, upper-level, and published academic writing. *Journal of English for Academic Purposes*, 26, 29-41. https://doi.org/10.1016/j.jeap.2017.01.005

Bahadofar, R., & Gholami, J. (2019). Types and functions of citations in master's theses across

disciplines and languages. *Discourse and Interaction*, 10(2), 27-45.

https://doi.org/10.5817/di2017-2-27

Bailey, S. (2014). *Academic writing: A handbook for international students*. Routledge.

https://doi.org/10.4324/9781315768960

Beatty, C. J., & Chan, M. J. (1984). Chinese scholars abroad: Changes in perceived academic

needs. *The ESP Journal*, 3(1), 53-59. https://doi.org/10.1016/0272-2380(84)90007-6

Belcher, D. D. (2006). English for specific purposes: Teaching to perceived needs and imagined

futures in worlds of work, study, and everyday life. *TESOL Quarterly*, 40(1), 133-156.

https://doi.org/10.2307/40264514

Benesch, S. (2009). Theorizing and practicing critical English for academic purposes. *Journal of

English for Academic Purposes*, 8(2), 81-85. https://doi.org/10.1016/j.jeap.2008.09.002

Berman, R., & Cheng, L. (2010). English academic language skills: Perceived difficulties by

undergraduate and graduate students, and their academic achievement. *Canadian

Journal of Applied Linguistics,* 4(1), 25-40.

https://journals.lib.unb.ca/index.php/CJAL/article/download/19830/21603

Bestgen, Y. (2018). Evaluating the frequency threshold for selecting lexical bundles by means

of an extension of the Fisher's exact test. *Corpora*, 13(2), 205-228.

https://doi.org/10.3366/cor.2018.0144

Bhatia, V. K. (2002). Professional discourse: Towards a multi-dimensional approach and shared practice. In C. C. Candlin (Ed.), *Research and practice in professional discourse,* (39–60). City University of Hong Kong Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.

https://doi.org/10.1093/applin/25.3.371

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

Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L2 thesis students writing the discussion section. *Journal of English for Academic Purposes*, 5(1), 4-18. https://doi.org/10.1016/j.jeap.2005.10.002

Björkman, B. (2011). English as a lingua franca in higher education: Implications for EAP. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos* (AELFE), (22), 79-100. https://dialnet.unirioja.es/descarga/articulo/3738644.pdf

Bolton, K., Nelson, G., & Hung, J. (2002). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7(2), 165-182. https://doi.org/10.1075/ijcl.7.2.02bol

Borg, E. (2000). Citation practices in academic writing. In P. Thompson (Ed.), *Patterns and perspectives: Insights into EAP writing practice,* (26-42). Centre for Applied Language

Studies.

https://www.researchgate.net/profile/Erik_Borg2/publication/313443669_Citation_pr
actices_in_academic_writing/links/5aa64e71aca272d448badeb1/Citation-practices-in-
academic-writing.pdf

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents.
ReCALL, 17(1), 47-64. https://doi.org/10.1017/s0958344005000510

Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific
Purposes*, 13(1), 47-59. https://doi.org/10.1016/0889-4906(94)90024-8

Bridgeman, B., & Carlson, S. B. (1984). Survey of academic writing tasks. *Written
Communication*, 1(2), 247-280. https://doi.org/10.1177/0741088384001002004

Brooke, M. (2013). Which research paradigm for TESOL?. *Theory & Practice in Language
Studies*, 3(3), 430-436. https://doi.org/10.4304/tpls.3.3.430-436

Browne, C., Culligan, B. & Phillips, J. (2013). The New Academic Word List. Retrieved February,
2020 from: http://www.newgeneralservicelist.org

Bruce, I. (2005). Syllabus design for general EAP writing courses: A cognitive approach. *Journal
of   English   for   Academic   Purposes*,   4(3),   239-256.
https://doi.org/10.1016/j.jeap.2005.03.001

Butler, C.S. (1990). Qualifications in science: Modal meanings in scientific texts. In W. Nash (Ed.)
The writing scholar: Studies in academic discourse, (137-170). Sage.

https://doi.org/10.2307/358234

Bychkovska, T., & Lee, J. J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes,* 30, 38-52. https://doi.org/10.1016/j.jeap.2017.10.008

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5(5), 31-64.

http://faculty.edfac.usyd.edu.au/projects/usp_in_TESOL/pdf/volume05/Article02.pdf

Cambridge English (2018). Cambridge English Corpus. Retrieved November, 2018, from: https://www.cambridge.org/gb/cambridgeenglish/better-learning/deeper-insights/linguistics-pedagogy/cambridge-english-corpus

Campion, M., & Elley, W. (1971). *An academic vocabulary list*. New Zealand Council For Educational Research.

Cao, Y., & Xiao, R. (2013). A multi-dimensional contrastive study of English abstracts by native and non-native writers. *Corpora*, 8(2), 209-234. https://doi.org/10.3366/cor.2013.0041

Carkin, S. (2005). English for academic purposes. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*, (85–98). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410612700

Casanave, C. P., & Hubbard, P. (1992). The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research. *English for Specific Purposes*, 11(1), 33-49. https://doi.org/10.1016/0889-4906(92)90005-u

Chan, V. (2001). Determining students' language needs in a tertiary setting. *English Teaching Forum*, 39(3), 16-27.

https://www.researchgate.net/profile/Victoria_Chan9/publication/265566434_Determining_Students'_Language_Needs_in_a_Tertiary_Setting/links/551bd8240cf2fe6cbf75f1d2/Determining-Students-Language-Needs-in-a-Tertiary-Setting.pdf

Chang, C. F., & Kuo, C. H. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, 30(3), 222-234. https://doi.org/10.1016/j.esp.2011.04.001

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310-331. https://doi.org/10.1016/j.esp.2005.05.003

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6(4), 289-302. https://doi.org/10.1016/j.jeap.2007.09.009

Charles, M. (2011). Using hands-on concordancing to teach rhetorical functions: Evaluation and implications for EAP writing classes. In A. Frankenberg-Garcia, L. Flowerdew, & G. Aston (Eds.), *New trends in corpora and language learning*, (26–43). Continuum. https://doi.org/10.5040/9781474211925.ch-002

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31(2), 93-102. https://doi.org/10.1016/j.esp.2011.12.003

305

Chen, Q., & Ge, G. C. (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. https://doi.org/10.1016/j.esp.2007.04.003

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30–49. http://dx.doi.org/10125/44213

Coates, J. (1983). *The Semantics of the modal auxiliaries*. Croom Helm.

Cobb, T. Compleat Web VP! [website]. Accessed Feb, 2020 at: http://www.lextutor.ca/vp/comp/

Conrad, S., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica* 20, 56–71.

https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1002&context=ling_fac

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.

https://doi.org/10.1016/j.esp.2003.12.001

Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In V. Cortes & E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (197-218). John Benjamins. https://doi.org/10.1075/scl.66.09cor

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. https://doi.org/10.2307/3587951

Coxhead, A. (2011). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355-362. https://doi.org/10.5054/tq.2011.254528

Coxhead, A. (2012). Academic vocabulary, writing and English for academic purposes: Perspectives from second language learners. *RELC Journal*, 43(1), 137-145. https://doi.org/10.1177/0033688212439323

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16(3), 129-147. https://doi.org/10.1016/j.jslw.2007.07.002

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes*, 16(4), 271-287. https://doi.org/10.1016/s0889-4906(97)00007-0

Crosthwaite, P., Cheung, L., & Jiang, F. K. (2017). Writing with attitude: Stance expression in learner and professional dentistry research reports. *English for Specific Purposes*, 46, 107-123. https://doi.org/10.1016/j.esp.2017.02.001

Daller, H., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In: H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge*, (150-164). Cambridge University Press. https://doi.org/10.1017/cbo9780511667268.011

Davis, M. (2013). The development of source use by international postgraduate students. *Journal of English for Academic Purposes*, 12(2), 125-135. https://doi.org/10.1016/j.jeap.2012.11.008

De Bartolo, A. M. (2018). Learners' awareness and attitude towards ELF. A pilot study in an

Italian University context. *Lingue e Linguaggi*, 26, 157-172.

https://doi.org/10.1285/i22390359v26p157

De Chazal, E, & Rogers, L. (2013). *Oxford EAP : A course in English for academic purposes.*

Intermediate-B1+. Oxford University Press.

Demir, C. (2018). Hedging and academic writing: an analysis of lexical hedges. *Journal of*

Language and Linguistic Studies, 14(4), 74-92.

https://dergipark.org.tr/en/download/article-file/650745

Diaz, I. (2015). Training in metacognitive strategies for students' vocabulary improvement by

using learning journals. *Profile Issues in Teachers Professional Development*, 17(1), 87-

102.

https://doi.org/10.15446/profile.v17n1.41632

Ding, H. (2007). Genre analysis of personal statements: Analysis of moves in application essays

to medical and dental schools. *English for Specific Purposes*, 26(3), 368-392.

https://doi.org/10.1016/j.esp.2006.09.004

Dong, Y. R. (1998). Non-native graduate students' thesis/dissertation writing in science: Self-

reports by students and their advisors from two US institutions. *English for Specific*

Purposes, 17(4), 369-390. https://doi.org/10.1016/s0889-4906(97)00054-9

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.

Dudley-Evans, T. and M. J. St John. (1998). *Developments in English for specific purposes*. Cambridge University Press.

Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, 35(3), 328-356. https://doi.org/10.1093/applin/amt016

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing?. *English for Specific Purposes*, 43, 49-61.

https://doi.org/10.1016/j.esp.2016.01.004

Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193.

https://doi.org/10.1093/applin/amv011

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations?. *IRAL- International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177. https://doi.org/10.1515/iral.2009.007

Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes*, 29, 21-42. https://doi.org/10.1016/j.jeap.2017.09.002

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6(1), 3-17.

https://doi.org/10.1016/j.jeap.2006.11.005

Evans, S., & Morrison, B. (2010). The first term at university: implications for EAP. *ELT journal*,

65(4), 387-397. https://doi.org/10.1093/elt/ccq072

Fazel, I., & Shi, L. (2015). Citation behaviours of graduate students in grant proposal writing.

Journal of English for Academic Purposes, 20, 203-214.

https://doi.org/10.1016/j.jeap.2015.10.002

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels

of L2 proficiency. *TESOL Quarterly*, 28(2), 414-420. https://doi.org/10.2307/3587446

Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong

Kong.   *Journal   of   Second   Language   Writing*,   8(3),   243-264.

https://doi.org/10.1016/s1060-3743(99)80116-7

Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the non-

native-English-speaking scholar. *TESOL Quarterly*, 34(1), 127-150.

https://doi.org/10.2307/3588099

Flowerdew,   J.   (2013).   *Academic   discourse.*   Routledge.

https://doi.org/10.4324/9781315838069

Flowerdew, L. (2000). Investigating referential and pragmatic errors in a learner corpus. In L.

Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective,*

(117–124). Peter Lang.

Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse*,

(95–114). Longman. https://doi.org/10.4324/9781315838069

Friginal, E., Li, M., & Weigle, S. C. (2014). Revisiting multiple profiles of learner compositions:

A comparison of highly rated L1 and L2 essays. *Journal of Second Language Writing*, 23,

1-16. https://doi.org/10.1016/j.jslw.2013.10.001

Gabrielatos, C. & McEnery, T. 2005. Epistemic modality in MA dissertations. In: P.A. Fuertes

Olivera (Ed.), *Lengua y sociedad: Investigaciones recientes en lingüística aplicada.*

Lingüística y Filología no. 61., (311-331). Universidad de Valladolid.

https://eprints.lancs.ac.uk/id/eprint/102/1/Epistemic_modality_in_MA_dissertations.p

df

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

305-327. https://doi.org/10.1093/applin/amt015

Gardner, S., & Nesi, H. (2012). A classification of genre families in university student writing.

Applied Linguistics, 34(1), 25-52. https://doi.org/10.1093/applin/ams024

Ghadessy, P. (1979). Frequency counts, word lists, material preparation: A new approach.

English Teaching Forum, 17, 24–27.

Gilquin, G. & Granger, S. (2011). From EFL to ESL: Evidence from the International Corpus

of Learner English. In M. Hundt & D. Mukherjee (Eds.), *Exploring second-language*

varieties of English and learner Englishes: Bridging a paradigm gap, (55–78).

John Benjamins. https://doi.org/10.1075/scl.44.04gra

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy.

Journal of English for Academic Purposes, 6(4), 319-335.

https://doi.org/10.1016/j.jeap.2007.09.007

Gilquin, G., & Paquot, M. (2007). Spoken features in learner academic writing: Identification, explanation and solution. In *Proceedings of the 4th corpus linguistics conference*, University of Birmingham (27-30).

http://ucrel.lancs.ac.uk/publications/cl2007/paper/204_Paper.pdf

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41-61. https://doi.org/10.1075/etc.1.1.05gil

Gollin-Kies, S. (2014). Methods reported in ESP research articles: A comparative survey of two leading journals. *English for Specific Purposes*, 36, 27-34. https://doi.org/10.1016/j.esp.2014.04.001

Gosden, H. (1992). Research writing and L2s: from the editors. *Journal of Second Language Writing*, 1(2), 123–139. https://doi.org/10.1016/1060-3743(92)90012-e

Granger, S., Dagneaux, E., & Meunier, F. (2002). *The International Corpus of Learner English. Handbook and CD-ROM*. Presses Universitaires de Louvain.

Granger, S., & Paquot, M. (2010) Customising a general EAP dictionary to learner needs. In S. Granger, & M. Paquot (Eds.). *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009*. Presses Universitaires de Louvain (83-86). Retrieved March, 2018, from:

https://s3.amazonaws.com/academia.edu.documents/4103268/elex_book_abstracts_final_new.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1521471915&Signature=2MvQ4YY22WJqv9Fe1L3eEj07fqo%3D&response-content-

disposition=inline%3B%20filename%3DIntroducing_the_MuLexFoR_A_Multilingual.pdf
#page=97

Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on computer*, (119-131). Longman.

https://doi.org/10.4324/9781315841342

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17-27. https://doi.org/10.1111/j.1467-971x.1996.tb00089.x

Graves, K., & Xu, S. (2000). *Designing language courses: A guide for teachers*. Heinle & Heinle.

Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education*, 14(1), 75-97. https://doi.org/10.1080/09695940701272880

Griffiths, C., & Soruç, A. (2019). Contextual differences regarding students' perceptions of English as a Lingua Franca according to subject major and nationality. *The Journal of Language Learning and Teaching*, 9(1), 53-69.

https://dergipark.org.tr/en/download/article-file/638291

Grix, J. (2004). *The foundations of research: A student's guide*. Macmillan International Higher Education.

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46(4), 643-679. https://doi.org/10.1111/j.1467-1770.1996.tb01355.x

Hamp-Lyons, L. (2011). English for academic purposes. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning,* (89-105). Routledge.

https://doi.org/10.4324/9780203836507

Hardy, J. A., & Friginal, E. (2016). Genre variation in student writing: A multi-dimensional analysis. *Journal of English for Academic Purposes*, 22, 119-131.

https://doi.org/10.1016/j.jeap.2016.03.002

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8(2), 183-207. https://doi.org/10.3366/cor.2013.0040

Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes*, 4(2), 149-161. https://doi.org/10.1016/j.jeap.2004.07.008

Harwood, N., & Hadley, G. (2004). Demystifying institutional practices: critical pragmatism and the teaching of academic writing. *English for Specific Purposes*, 23(4), 355-377.

https://doi.org/10.1016/j.esp.2003.08.001

Harwood, N., & Petrić, B. (2012). Performance in the citing behavior of two student writers. *Written Communication*, 29(1), 55-103. https://doi.org/10.1177/0741088311424133

Hawkey, R. A. (2006) Impact theory and practice: studies of the IELTS test and Progetto Lingue 2000. *Studies in language testing 24.* Cambridge ESOL/Cambridge University Press.

https://ugeb.pw/buxi-kykav-p-sixu-m.pdf

Heatley, A., & Nation, P. (2002). Range [Computer software]. Wellington, New Zealand: Victoria University of Wellington. (Available from: https://www.victoria.ac.nz/lals/resources/range)

Helmer, K. A. (2013). Critical English for academic purposes: Building on learner, teacher, and program strengths. *Journal of English for Academic Purposes*, 12(4), 273-287. https://doi.org/10.1016/j.jeap.2013.08.003

HESA - Higher Education Statistics Agency (2018). HE Student Enrolments by Domicile. Retrieved February, 2018, from: https://www.hesa.ac.uk/data-and-analysis/sfr247/figure-8

Hewings, A., Lillis, T., & Vladimirou, D. (2010). Who's citing whose writings? A corpus based study of citations as interpersonal resource in English medium national and English medium international journals. *Journal of English for Academic Purposes*, 9(2), 102-115. https://doi.org/10.1016/j.jeap.2010.02.005

Hewings, M. (2012). *Cambridge Academic English Student's Book, Upper Intermediate*. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781139383370

Higginbotham, G., & Reid, J. (2019). The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes*, 37, 127-140. https://doi.org/10.1016/j.jeap.2018.12.002

Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, 12(2), 111-132.

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.9191&rep=rep1&type=pdf

Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37(2), 275-301. https://doi.org/10.2307/3588505

Hinkel, E. (2005). Hedging, inflating, and persuading in L2 academic writing. *Applied Language Learning*, 15(1/2), 29-53.

https://books.google.co.uk/books?hl=en&lr=&id=VzZ8ZHWyr7YC&oi=fnd&pg=PA29&dq=Hedging,+inflating,+and+persuading+in+L2+academic+writing&ots=7DCUD4XmCM&sig=X4ZHJ25uPmdNgpFEgzYgj6wSOXw#v=onepage&q=Hedging%2C%20inflating%2C%20and%20persuading%20in%20L2%20academic%20writing&f=false

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41(4), 667-683. https://doi.org/10.1016/j.pragma.2008.09.029

Hirvela, A., & Du, Q. (2013). "Why am I paraphrasing?": Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes*, 12(2), 87-98. https://doi.org/10.1016/j.jeap.2012.11.005

Hu, G. (2007). Developing an EAP writing course for Chinese ESL students. *RELC Journal*, 38(1), 67-86. https://doi.org/10.1177/0033688206076160

Hu, G., & Wang, G. (2014). Disciplinary and ethnolinguistic influences on citation in research articles. *Journal of English for Academic Purposes*, 14, 14-28.

https://doi.org/10.1016/j.jeap.2013.11.001

Hu, Z., Brown, D., & Brown, L. (1982). Some linguistic differences in the written English of Chinese and Australian students. *Language Learning and Communication*, 1(1), 39-49

Huddleston, R. (1971). *The Sentence in written English: A Syntactic study based on an analysis of scientific texts*. Cambridge University Press.

Hundt, M., Schneider, G., & Seoane, E. (2016). The use of the be-passive in academic Englishes: local versus global usage in an international language. *Corpora*, 11(1), 29-61.

https://doi.org/10.3366/cor.2016.0084

Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge University Press.

https://doi.org/10.1017/cbo9780511733031

Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, 13(3), 239-256. https://doi.org/10.1016/0889-4906(94)90004-3

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 17(4), 433-454. https://doi.org/10.1093/applin/17.4.433

Hyland, K. (1998), *Hedging in scientific research articles.* John Benjamins. https://doi.org/10.1075/pbns.54

Hyland, K. (1999a) Disciplinary discourses: Writer stance in research articles. In C. Candlin and K. Hyland (Eds.) *Writing: Texts: Processes and practices*, (99–121). Longman.

https://doi.org/10.4324/9781315840390-6

Hyland, K. (1999b). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3), 341-367. https://doi.org/10.1093/applin/20.3.341

Hyland, K. (2000). *Disciplinary discourse: Social interactions in academic writing*. Longman.

https://doi.org/10.3998/mpub.6719

Hyland, K. (2002a). Specificity revisited: how far should we go now?. *English for Specific Purposes*, 21(4), 385-395. https://doi.org/10.1016/s0889-4906(01)00028-x

Hyland, K. (2002b). Activity and evaluation: reporting practices in academic writing. In J. Flowerdew (Ed.) *Academic discourse,* (115-130). Longman.

https://doi.org/10.4324/9781315838069

Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13(2), 133-151. https://doi.org/10.1016/j.jslw.2004.02.001

Hyland, K. (2005) *Metadiscourse*. Continuum.

Hyland, K. (2006). Disciplinary differences: Language variation in academic discourses. In K. Hyland, & M. Bondi (Eds.) *Academic discourse across disciplines*, (17-45). Peter Lang.

https://doi.org/10.3726/978-3-0351-0446-2

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

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English for Academic Purposes*, 1(1), 1-12. https://doi.org/10.1016/s1475-1585(02)00002-4

Hyland, K., & Jiang, K. F. (2018). Academic lexical bundles: how are they changing? *International Journal of Corpus Linguistics*, 23(4), 383–407. https://doi.org/10.1075/ijcl.17080.hyl

Hyland, K., & Jiang, F. (2019). Points of reference: Changing patterns of academic citation. *Applied Linguistics*, 40(1), 64-85. https://doi.org/10.1093/applin/amx012

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183-205. https://doi.org/10.1016/s1060-3743(97)90033-3

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"?. *TESOL Quarterly*, 41(2), 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies*, 9(2), 111-129.

https://revistas.um.es/ijes/article/download/90781/87581

Ivanič, R., & Camps, D. (2001). I am how I sound: Voice as self-representation in L2 writing. Journal of second language writing, 10(1-2), 3-33.

https://doi.org/10.1016/s1060-3743(01)00034-0

Jalilifar, A. R. (2012a). Academic attribution: Citation analysis in master's theses and research articles in applied linguistics. *International Journal of Applied Linguistics*, 22(1), 23-41.

https://doi.org/10.1111/j.1473-4192.2011.00291.x

Jalilifar, A. R. (2012b). Research article introductions: Sub-disciplinary variations in applied linguistics. *Journal of Teaching Language Skills*, 29(2), 29-55.

https://www.sid.ir/EN/VEWSSID/J_pdf/13112010600402.pdf

Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12(4), 377-403. https://doi.org/10.1016/j.jslw.2003.09.001

Jasso-Aguilar, R. (1999). Sources, methods and triangulation in needs analysis: A critical perspective in a case study of Waikiki hotel maids. *English for Specific Purposes*, 18(1), 27-46. https://doi.org/10.1016/s0889-4906(97)00048-3

Johansson, S. (1991) Computer corpora in English language research. In S. Johansson and A.-B. Stenström (Eds.), *English computer corpora: Selected papers and research guide*, 3-6. Mouton de Gruyter. https://doi.org/10.1515/9783110865967.3

Jomaa, N. J. (2019). Perspectives of EFL doctoral students on challenges of citations in academic writing. *Malaysian Journal of Learning and Instruction*, 14(2), 177-209. https://doi.org/10.32890/mjli2017.14.2.7

Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers.* Cambridge University Press.

Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, 1(1), 69-78. https://doi.org/10.1016/s1475-1585(02)00004-8

Kafes, H. (2017). Citation practices among novice and expert academic writers. *Egitim ve Bilim*, 42(192), 441-462. https://doi.org/10.15390/eb.2017.6317

Kanoksilapatham, B. (2015). Distinguishing textual features characterizing structural variation in research articles across three engineering sub-discipline corpora. *English for Specific Purposes*, 37, 74-86. https://doi.org/10.1016/j.esp.2014.06.008

Kaypak, E., & Ortaçtepe, D. (2014). Language learner beliefs and study abroad: A study on English as a lingua franca (ELF). *System*, 42, 355-367. https://doi.org/10.1016/j.system.2014.01.005

Kennedy, G. (1998). *An introduction to corpus linguistics*. Routledge. https://doi.org/10.4324/9781315843674

Kim, L. C., & Lim, J. M. H. (2015). Hedging in academic writing-a pedagogically-motivated qualitative study. *Procedia-Social and Behavioral Sciences*, 197, 600-607. https://doi.org/10.1016/j.sbspro.2015.07.200

Klimova, B. F. (2012). Changes in the notion of academic writing. *Procedia-Social and Behavioural Sciences*, 47, 311-315. https://doi.org/10.1016/j.sbspro.2012.06.656

Krishnamurthy, R., & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for Academic Purposes*, 6(4), 356-373. https://doi.org/10.1016/j.jeap.2007.09.003

Kwan, B. S. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25(1), 30-55. https://doi.org/10.1016/j.esp.2005.06.001

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic

    disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 131-140.

    https://doi.org/10.17509/ijal.v7i1.6866

Le Ha, P. (2009). Strategic, passionate, but academic: Am I allowed in my writing?. *Journal of*

    English for Academic Purposes, 8(2), 134-146.

    https://doi.org/10.1016/j.jeap.2008.09.003

Lea, M. R. (2004). Academic literacies: A pedagogy for course design. *Studies in Higher*

    Education, 29(6), 739-756. https://doi.org/10.1080/0307507042000287230

Lea, M. R., & Stierer, B. (2000). *Student writing in higher education: New contexts*. Open

    University Press/Society for Research into Higher Education.

    http://oro.open.ac.uk/21798/2/lea_book.pdf

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies

    approach. *Studies in Higher Education*, 23(2), 157-172.

    https://doi.org/10.1080/03075079812331380364

Lee, D., & Swales, J. (2006). A corpus-based EAP course for L2 doctoral students: Moving from

    available specialized corpora to self-compiled corpora. *English for Specific Purposes*,

    25(1), 56-75. https://doi.org/10.1016/j.esp.2005.02.010

Lee, J. J., Hitchcock, C., & Casal, J. E. (2018). Citation practices of L2 university students in first-

    year writing: Form, function, and stance. *Journal of English for Academic Purposes*, 33,

    1-11. https://doi.org/10.1016/j.jeap.2018.01.001

Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101.

https://doi.org/10.2307/3587199

Leki, I., & Carson, J. (1997). "Completely different worlds": EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly*, 31(1), 39-69.

https://doi.org/10.2307/3587974

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.

https://doi.org/10.1016/j.jslw.2009.02.001

Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes*, 11(4), 345-356. https://doi.org/10.1016/j.jeap.2012.07.004

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35.

https://doi.org/10.1016/j.esp.2011.07.002

Liyanage, I., & Birch, G. (2001). English for general academic purposes: Catering to discipline-specific needs. *Queensland Journal of Educational Research*, 17(1), 48-67. https://www.researchgate.net/profile/Indika_Liyanage/publication/29454991_English_for_General_Academic_PurposesCatering_for_Discipline_Specific_Needs/links/00463526b8a0b00dad000000/English-for-General-Academic-PurposesCatering-for-Discipline-Specific-Needs.pdf

Loi, C. K., & Lim, J. M. H. (2019). Hedging in the discussion sections of English and Malay Educational Research Articles. *GEMA Online® Journal of Language Studies*, 19(1), 36-61.

https://doi.org/10.17576/gema-2019-1901-03

Long, M. H. (2005). Methodological issues in learner needs analysis. In M. H. Long (Ed.),

Second language needs analysis, (19–76). Cambridge University Press.

https://doi.org/10.1017/cbo9780511667299.002

Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer,* (53-66). Longman.

https://doi.org/10.4324/9781315841342

Lynn, R. W. (1973). Preparing word lists a suggested method. *RELC Journal*, 4(1), 25–28.

https://doi.org/10.1177/003368827300400103

Lyons, J. (1977) *Semantics*. Cambridge University Press.

Mack, L. (2010). The philosophical underpinnings of educational research. *Polyglossia*, 19, 5-11.

https://pdfs.semanticscholar.org/9f5f/c9abeb1467d2c0242ddbf6128120c9bcf324.pdf

Maley, A. (2009). ELF: a teacher's perspective. *Language and Intercultural Communication*, 9(3), 187-200. https://doi.org/10.1080/14708470902748848

Malmström, H., Pecorari, D., & Shaw, P. (2018). Words for what? Contrasting university students' receptive and productive academic vocabulary needs. *English for Specific Purposes*, 50, 28-39. https://doi.org/10.1016/j.esp.2017.11.002

Mansourizadeh, K., & Ahmad, U. K. (2011). Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes*, 10(3), 152-161. https://doi.org/10.1016/j.jeap.2011.03.004

Martínez, I. A. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing*, 14(3), 174-190. https://doi.org/10.1016/j.jslw.2005.06.001

McCarten, J. (2007). Teaching vocabulary. Lessons from the corpus, lessons for the classroom. Cambridge University Press [online].

https://www.academia.edu/download/58393723/McCarten-Teaching-Vocabulary.pdf

McDonough, K., Neumann, H., & Hubert-Smith, N. (2018). How accurately do English for academic purposes students use Academic Word List words?. *BC TEAL*, 3(1), 77-89. https://spectrum.library.concordia.ca/985472/1/BC_teal_accuracy_of_AWL.pdf

McEnery, A. M., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.

McEnery, T., & Gabrielatos, C. (2006). English corpus linguistics. In B. Aarts and A. McMahon (Eds.) *The handbook of English linguistics*, 33-71. Blackwell.

https://doi.org/10.1002/9780470753002.ch3

McGrath, L. (2016). Self-mentions in anthropology and history research articles: Variation between and within disciplines. *Journal of English for Academic Purposes*, 21, 86-98. https://doi.org/10.1016/j.jeap.2015.11.004

Milton, J. (1998). Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment. In S. Granger (Ed.), *Learner English on computer*, 186-198. Longman. https://doi.org/10.4324/9781315841342

Mizumoto, A., & Takeuchi, O. (2008). Exploring the driving forces behind TOEIC scores: Focusing on vocabulary learning strategies, motivation, and study time. *JACET Journal*, 46, 17-32.

http://www.academia.edu/download/36712621/110007137438.pdf

Mokhtar, A. A., Rawian, R. M., Yahaya, M. F., Abdullah, A., & Mohamed, A. R. (2017). Vocabulary learning strategies of adult ESL learners. *The English Teacher* XXXVIII, 133-145.

https://journals.melta.org.my/index.php/tet/article/download/290/187

Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4(1), 43-66. https://doi.org/10.1016/j.jeap.2004.02.001

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25(2), 235–256. https://doi.org/10.1016/j.esp.2005.05.002

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2–13. https://doi.org/10.2167/illt039.0

Nesi, H., Gardner, S., Forsyth, R., Hindle, D., Wickens, P., Ebeling, S., et al. (2005). Towards the compilation of a corpus of assessed student writing: An account of work in progress. In P. Danielsson, & M. Wagenmakers (Eds.), *Proceedings from the Corpus Linguistics*

Conference series. Birmingham, University of Birmingham. Retrieved August, 2018, from:

https://warwick.ac.uk/fac/soc/al/research/collections/bawe/papers/towards_compilat

ion.pdf

Nesi, H., Sharpling, G., & Ganobcsik-Williams, L. (2004). Student papers across the curriculum:

Designing and developing a corpus of British student writing. *Computers and

Composition*, 21(4), 439-450. https://doi.org/10.1016/s8755-4615(04)00040-4

Nguyen, T. T.L., & Pramoolsook, I. (2015). Reporting verbs in literature review chapters of

TESOL master's theses written by Vietnamese postgraduates. *ESP Today-Journal OF

English for Specific Purposes at Tertiary Level*, 3(2), 196-215.

Nosratinia, M., & Zaker, A. (2015). Boosting autonomous foreign language learning:

Scrutinizing the role of creativity, critical thinking, and vocabulary learning strategies.

International Journal of Applied Linguistics and English Literature, 4(4), 86-97.

https://doi.org/10.7575/aiac.ijalel.v.4n.4p.86

Okamura, A. (2008). Use of citation forms in academic texts by writers in the L1 & L2 context.

The Economic Journal of Takasaki City University of Economics, 51(1), 29-44.

https://pdfs.semanticscholar.org/692a/0951b5dd4adc8fe5d7fd46816990a0314070.pdf

Oxford, R.L. (2002). Language learning strategies in a nutshell: Update and ESL suggestions. In

J.C. Richards, W.A. Renandya, (Eds.), *Methodology in Language Teaching*, 124-132.

Cambridge University Press.

Ozturk, I. (2007). The textual organization of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, 26 (1), 25-38. https://doi.org/10.1016/j.esp.2005.12.003

Paltridge, B. (2002). Thesis and dissertation writing: an examination of published advice and actual practice. *English for Specific Purposes*, 21(2), 125-143. https://doi.org/10.1016/s0889-4906(00)00025-9

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. https://doi.org/10.1016/j.jeap.2015.11.003

Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal*, 47(1), 10-11. https://pdfs.semanticscholar.org/b05f/336c4cea9f2e74579b38eab3bb658485ce45.pdf

Paquot, M. (2005). EAP vocabulary in learner corpora. A cross-linguistic perspective. In C. Cosme, C. Gouverneur, F. Meunier, & M. Paquot (Eds.), *Proceedings of the Phraseology 2005 Conference*, (323-326). https://sites.uclouvain.be/cecl/projects/Downloads/Paquot%202005.doc

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], (127–140). Frankfurt am Main, Germany. https://dial.uclouvain.be/downloader/downloader.php?pid=boreal:76045&datastream=PDF_01

Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, 30(4), 479-497. https://doi.org/10.1016/s0346-251x(02)00050-7

Pecorari, D. (2006). Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes*, 25(1), 4-29.

https://doi.org/10.1016/j.esp.2005.04.004

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94.

https://doi.org/10.1016/j.jeap.2014.01.002

Petchkij, W. (2019). Explicit teaching of hedges: Bringing hedging in academic writing into the Thai EFL classroom. *Electronic Journal of Foreign Language Teaching*, 16(1), 95-113.

http://e-flt.nus.edu.sg/v16n12019/petchkij.pdf

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes*, 20, 28-39. https://doi.org/10.1016/j.jeap.2015.04.002

Petrić, B. (2007). Rhetorical functions of citations in high-and low-rated master's theses. *Journal of English for Academic Purposes,* 6(3), 238-253.

https://doi.org/10.1016/j.jeap.2007.09.002

Petrić, B. (2012). Legitimate textual borrowing: Direct quotation in L2 student writing. *Journal of Second Language Writing*, 21(2), 102-117. https://doi.org/10.1016/j.jslw.2012.03.005

Pickard, V. (1995). Citing previous writers: What can we say instead of' 'Say'? *Hong Kong Papers in Linguistics and Language Teaching*, 18, 89-102.

https://files.eric.ed.gov/fulltext/ED390262.pdf

Poos, D., & Simpson, R. (2002). Cross-disciplinary comparisons of hedging. Some findings from

the MICASE. In: R. Reppen; S.M. Fitzmaurice & D. Biber (eds.) *Using corpora to explore*

linguistic variation, (3-24). John Benjamins. https://doi.org/10.1075/scl.9.03poo

Praninskas, J. (1972). *American university word list*. Longman.

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in*

Accounting & Management, 8(3), 238-264.

https://doi.org/10.1108/11766091111162070

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*,

9(2), 109-121. https://doi.org/10.1016/0889-4906(90)90002-t

Ren, H., & Li, Y. (2011). A comparison study on the rhetorical moves of abstracts in published

research articles and Master's foreign-language theses. *English Language Teaching*, 4(1),

162-166. https://doi.org/10.5539/elt.v4n1p162

Robson, C. (2011) *Real world research: a resource for users of social research methods in*

applied settings. Wiley.

Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): the

design, compilation and genre classification of the Michigan Corpus of Upper-level

Student Papers (MICUSP). *Corpora*, 6(2), 159-177.

https://doi.org/10.3366/cor.2011.0011

Römer, U., & Wulff, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research*, 2(2), 99-127. https://doi.org/10.17239/jowr-2010.02.02.2

Ryoo, Y. S. (2018). Comparing lexical diversity and lexical sophistication in Korean EFL writing: Topic and text length. *Multimedia-Assisted Language Learning*, 21(3), 63-87. http://journal.kamall.or.kr/wp-content/uploads/2018/10/21_3_03_Ryoo.pdf

Samraj, B. (2013). Form and function of citations in discussion sections of master's theses and research articles. *Journal of English for Academic Purposes*, 12(4), 299-310. https://doi.org/10.1016/j.jeap.2013.09.001

Schmitt, N., & Schmitt, D. (1995). Vocabulary notebooks: Theoretical underpinnings and practical implications. *ELT Journal*, 49(2), 133-143. https://doi.org/10.1093/elt/49.2.133

Scott, M., (2019), WordSmith Tools (Version 7) [computer software]. Stroud, England. Available from: https://www.lexically.net/wordsmith/downloads/

Shahriari, H. (2017). Comparing lexical bundles across the introduction, method and results sections of the research article. *Corpora*, 12(1), 1-22. https://doi.org/10.3366/cor.2017.0107

Shi, L. (2010). Textual appropriation and citing behaviors of university undergraduates. *Applied Linguistics*, 31(1), 1-24. https://doi.org/10.1093/applin/amn045

Shi, L. (2012). Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing*, 21(2), 134-148. https://doi.org/10.1016/j.jslw.2012.03.003

Shing, S., & Sim, T. (2011). EAP needs analysis in higher education: Significance and future direction. *English for Specific Purposes World*, 33(11), 1-11. https://www.researchgate.net/profile/Tam_Sim/publication/267411299_EAP_Needs_Analysis_in_Higher_Education_Significance_and_Future_Direction/links/555bf18008ae8f66f3adbccb.pdf

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). The Michigan corpus of academic spoken English. *The Regents of the University of Michigan*.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. https://doi.org/10.1093/applin/amp058

Steinman, L. (2003). Cultural collisions in L2 academic writing. *TESL Canada Journal*, 20(2), 80-91. https://doi.org/10.18806/tesl.v20i2.950

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swan, M. (2017). EFL, ELF, and the question of accuracy. *ELT Journal*, 71(4), 511-515.

https://doi.org/10.1093/elt/ccx031

Takimoto, M. (2015). A corpus-based analysis of hedges and boosters in English academic articles. *Indonesian Journal of Applied Linguistics*, *5*(1), 95-105.

https://doi.org/10.17509/ijal.v5i1.836

Tanaka, J., & Gilliland, B. (2017). Critical thinking instruction in English for academic purposes writing courses: A dialectical thinking approach. *TESOL Journal*, 8(3), 657-674.

https://doi.org/10.1002/tesj.291

Taş, E. E. I. (2010). "In this paper I will discuss…": Current trends in academic writing. *Procedia-Social and Behavioural Sciences*, 3, 121-126.

https://doi.org/10.1016/j.sbspro.2010.07.022

Terraschke, A., & Wahid, R. (2011). The impact of EAP study on the academic experiences of international postgraduate students in Australia. *Journal of English for Academic Purposes*, 10(3), 173-182. https://doi.org/10.1016/j.jeap.2011.05.003

Thompson, G., & Yiyun, Y. (1991). Evaluating in the reporting verbs used in academic papers. *Applied Linguistics*, 12(4), 365–382. https://doi.org/10.1093/applin/12.4.365

Thompson, P. (2006) 'Assessing the contribution of corpora to EAP practice'. In Z. Kantaridou, I. Papadopoulou & I. Mahili (Eds.) *Motivation in learning language for specific and academic purposes,* (1-19). University of Macedonia.

http://www.reading.ac.uk/internal/appling/thompson_macedonia.pdf

Tribble, C., & Wingate, U. (2013). From text to corpus – A genre-based approach to academic literacy instruction. *System*, 41(2), 307-321.

https://doi.org/10.1016/j.system.2013.03.001

University of Nottingham. (May, 2020). Centre for English Language Education – Which pre-sessional is right for you? https://www.nottingham.ac.uk/cele/presessional/which-course/index.aspx

Uysal, H. H. (2012). Argumentation across L1 and L2 writing: Exploring cultural influences and transfer issues. *Vigo International Journal of Linguistics*, 9, 133-159.

http://revistas.webs.uvigo.es/index.php/vial/article/download/53/53

Vandenhoek, T. (2018). Epistemic markers in NS and NNS academic writing. *Journal of Academic Writing*, *8*(1), 72-91. https://doi.org/10.18552/joaw.v8i1.359

Varley, S. (2009). I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133–152. https://doi.org/10.1080/09588220902778294

Varttala, T. (1999). Remarks on the communicative functions of hedging in popular scientific and specialist research articles on medicine. *English for Specific Purposes*, 18(2), 177-200. https://doi.org/10.1016/s0889-4906(98)00007-6

Varttala, T. (2001). *Hedging in scientifically oriented discourse. Exploring variation according to discipline and intended audience*. Tampere University Press. Retrieved April, 2018, from:

https://tampub.uta.fi/bitstream/handle/10024/67148/95144-5195-3.pdf?sequence=1

Vázquez, I., & Giner, D. (2008). Beyond mood and modality: Epistemic modality markers as hedges in research articles. A cross-disciplinary study. *Revista Alicantina de Estudios Ingleses* 21,171-190. https://doi.org/10.14198/raei.2008.21.10

Vold, E. T. (2006). Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics*, 16(1), 61-87.

https://doi.org/10.1111/j.1473-4192.2006.00106.x

Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33-41.

https://doi.org/10.1016/j.esp.2008.08.003

Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442-458. https://doi.org/10.1016/j.esp.2008.05.003

Wang, Y. L. (2014). A corpus-based contrastive study on academic vocabulary of RA abstracts in applied linguistics. *Applied Mechanics and Materials* 631, 1442-1446.

https://doi.org/10.4028/www.scientific.net/amm.631-632.1442

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28, 170–182.

https://doi.org/10.1016/j.esp.2009.04.001

Weber, W. (2004). The rhetoric of positivism versus interpretivism. *MIS Quarterly*, 28(1), iii-xii.

https://doi.org/10.2307/25148621

West, M. (1953). *A general service list of English words*. Longman.

Wette, R. (2010). Evaluating student learning in a university-level EAP unit on writing using sources. *Journal of Second Language Writing*, 19(3), 158-177.

https://doi.org/10.1016/j.jslw.2010.06.002

Wingate, U., & Tribble, C. (2011). The best of both worlds? Towards an English for Academic Purposes/Academic Literacies writing pedagogy. *Studies in Higher Education*, 37(4), 481-495. https://doi.org/10.1080/03075079.2010.525630

Woodward-Kron, R. (2008). More than just jargon–the nature and role of specialist language in learning disciplinary knowledge. *Journal of English for Academic Purposes*, 7(4), 234-249. https://doi.org/10.1016/j.jeap.2008.10.004

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28. https://doi.org/10.1016/s0271-5309(99)00015-4

Wulff, S., & Römer, U. (2009). Becoming a proficient academic writer: Shifting lexical preferences in the use of the progressive. *Corpora*, 4(2), 115-133. https://doi.org/10.3366/e1749503209000276

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215–219.

Yakhontova, T. (2006). Cultural and disciplinary variation in academic discourse: The issue of influencing factors. *Journal of English for Academic Purposes*, 5(2), 153-167. https://doi.org/10.1016/j.jeap.2006.03.002

Yilmaz, K. (2013). Comparison of quantitative and qualitative research traditions: Epistemological, theoretical, and methodological differences. *European Journal of Education*, 48(2), 311-325. https://doi.org/10.1111/ejed.12014

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12, 31–48. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44142/12_02_yoon.pdf

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283. https://doi.org/10.1016/j.jslw.2004.06.002

Yüksel, H. G., & Kavanoz, S. (2018). Dimension of experience: Metadiscourse in the texts of novice non-native, novice native and expert native speaker. *Advances in Language and Literary Studies*, 9(3), 104-112. https://doi.org/10.7575/aiac.alls.v.9n.3p.104

Zhang, Y., & Du, X. (2018). Chinese university students' and teachers' perceptions of and attitudes towards ELF. *Journal of Pan-Pacific Association of Applied Linguistics*, 22(2), 1-25. https://doi.org/10.25256/paal.22.2.1

Zhao, C. G., & Llosa, L. (2008). Voice in high-stakes L1 academic writing assessment: Implications for L2 writing instruction. *Assessing Writing*, 13(3), 153-170. https://doi.org/10.1016/j.asw.2008.10.003

Appendices

Appendix A Ethical Documentation

A1 Risk Assessment

Manchester
Metropolitan
University

RISK ASSESSMENT

FACULTY/DEPARTMENT	BUILDING
Art and Humanities/Languages, Linguistics and TESOL	Geoffrey Manton

1) ACTIVITY

PhD Research – Interviews/questionnaires conducted with MMU students, analysis of academic corpus.

2) PERSONS AT RISK

Study Participants

3A) HAZARDS

Data security

3B) Hazard Rating	Low ☒	Med ☐	High ☐

4A) RISK CONTROL METHODS

Participants' data will be anonymised, and all data collected will be stored on a secure personal laptop and encrypted external drive, both of which will be accessible solely by the researcher. Any information pertaining to the identities or personal contact details of the participants will be destroyed once the study is complete.

4B) Hazard Rating with control methods	Low ☒	Med ☐	High ☐

5) FURTHER ACTION REQUIRED

None

NAME AND TITLE OF ASSESSOR	SIGNATURE	DATE
Andrew Coates		26/10/17
		REVIEW DATE
SIGNATURE OF DEAN/HEAD OF DEPT.		DATE 27/10/17.
		REVIEW DATE

A2 Ethics Approval

Manchester Metropolitan
University

Faculty of Arts and Humanities
Research and Knowledge Exchange

Manchester Metropolitan University, Room 123, Geoffrey Manton Building, Rosamund Street West, Off Oxford Road, Manchester, M15 6LL, UK
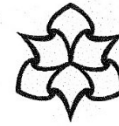
+44 (0)161 247 6673

8 February 2018

Dear Dr Drummond,

Re: Ethics Checklist

Project Title: The potential for a genre and corpus analysis of university students' academic writing to inform content in a university pre-sessional course.

I am pleased to inform you that Andrew Coates' Ethics Checklist has been approved unconditionally. This part of the research can now begin.

Yours sincerely

KWalthall.

Katherine Walthall
Research Group Officer

Tel: +44 (0)161 247 6673
Email: k.walthall@mmu.acuk
Research and Knowledge Exchange Office
Room 123 Geoffrey Manton Building

cc.   Director of Studies
       Departmental Research Degrees Co-ordinator
       Research Degrees Administrator
       Applicant

INVESTORS IN PEOPLE | Champion

www.mmu.ac.uk

A3 Ethics Amendments Approval

Manchester Metropolitan
University

Faculty of Arts and Humanities
Research and Knowledge Exchange

Manchester Metropolitan University, Room 123, Geoffrey Manton Building, Rosamund Street West, Off Oxford Road, Manchester, M15 6LL, UK

+44 (0)161 247 6673

Dr Rob Drummond
Department of Languages and Info Comms

26 April 2018

Dear Dr Rob Drummond

Re: Ethics Checklist Amendment

Project Title: The potential for a genre and corpus analysis of university students' academic writing to inform content in a university pre-sessional course [Andrew Coates]

I am pleased to inform you that the amendment to the above titled Ethics Checklist has been approved unconditionally. This part of the research can now begin.

Yours sincerely

KAWatthall.

Katherine Walthall
Research Group Officer

Tel: +44 (0)161 247 6673
Email: k.walthall@mmu.acuk
Research and Knowledge Exchange Office
Room 123 Geoffrey Manton Building

cc.   Departmental Research Degrees Co-ordinator
       Research Degrees Administrator
       Applicant

INVESTORS IN PEOPLE | Champion

www.mmu.ac.uk

A4 Confidentiality Letter Agreement

Manchester Metropolitan
University

All Saints Building,
All Saints, Manchester,
M15 6BH, UK

+44 (0)161 247 4696

Andrew Coates

By email: r.drummond@mmu.ac.uk

29 June 2018

Dear Andrew,

Confidentiality Letter Agreement

1 Disclosure

1.1 We, Manchester Metropolitan University (Manchester Met) understand that you, Andrew Coates (Recipient) have requested access to Moodle for the specific purpose of downloading coursework submitted by students on the Manchester Met's Postgraduate Linguistics Network who have agreed to take part in your PHD research. You have obtained from specific students their express consent to access their coursework (submitted through Moodle) and your aim is to analyse student textual use in coursework. You are a PHD student as well as an Associate Lecturer for the Manchester Met and you are currently subject to ongoing obligations of confidentiality as a result of such statuses.

1.2 You understand that:

    1.2.1   You must keep a written record for every consent obtained and when it is withdrawn (where relevant);

    1.2.2   You will only access coursework where you have written authorisation from the student;

    1.2.3   you must not access information on Moodle about students (including coursework) who have not expressly agreed to take part in your PHD research;

    1.2.4   you will immediately stop accessing student records where a student has informed you they no longer wish to participate; and

    1.2.5   you will not use (for any reason) information available on Moodle apart for coursework where you have received express consent.

1.3 In this letter agreement, Confidential Information means all confidential or proprietary information (however recorded or preserved) that is available to the Recipient through Moodle and which is not student coursework submitted by students who have consented to participate into the PHD research unless consent has been withdrawn, by Manchester Met to the Recipient.

1.4 For avoidance of doubt, all personal information (such as name, email address and student number) relating to students, whether or not they have consented to take part in the PHD research are Confidential Information.

INVESTORS IN PEOPLE | Champion
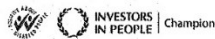
www.mmu.ac.uk
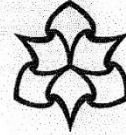
1.5 In consideration of Manchester Met granting access to Confidential Information to the Recipient, the Recipient undertakes to Manchester Met that it shall:

    1.5.1   keep the Confidential Information secret and confidential;

    1.5.2   not use or exploit the Confidential Information in any way; and

    1.5.3   not make any disclosure of the Confidential Information.

2 Limitation on obligations

The obligations set out in Paragraph 1 shall not apply, or shall cease to apply, to Confidential Information which the Recipient can show to Manchester Met's reasonable satisfaction:

2.1 that it is, or becomes, generally available to the public, other than as a direct or indirect result of the information being disclosed by the Recipient in breach of this letter agreement; or

2.2 was already lawfully known to the Recipient before it was disclosed by Manchester Met; or

2.3 has been received by the Recipient from a third party source that is not connected with Manchester Met and that such source was not under any obligation of confidence in respect of that information.

3 Term and termination

3.1 This letter agreement will terminate on such date Manchester Met will remove access to Moodle or until the date of submission of his findings/research by the Recipient.

3.2 The obligations of the Recipient shall continue for a period of one (1) year from the termination of this letter agreement.

3.3 The termination of this letter agreement shall not affect any accrued rights or remedies to which either party is entitled.

4 Governing law and jurisdiction

4.1 This letter agreement and any dispute or claim (including non-contractual disputes or claims) arising out of or in connection with it or its subject matter or formation shall be governed by and construed in accordance with the law of England and Wales.

4.2 Each party irrevocably agrees that the courts of England and Wales shall have exclusive jurisdiction to settle any dispute or claim (including non-contractual disputes or claims) arising out of or in connection with this letter agreement or its subject matter or formation.

Please sign and return a copy of this letter agreement if you agree to its terms.
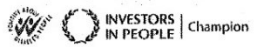
Yours faithfully,

Signed by
for and on behalf of Manchester Metropolitan University
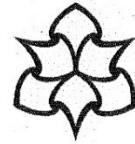
We acknowledge receipt and agree to the terms of this letter agreement:
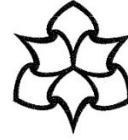..

Signed by Andrew Coates
Date ....02/07/18..................................

Manchester Metropolitan
University

All Saints Building,
All Saints, Manchester,
M15 6BH, UK

+44 (0)161 247 4696

Andrew Coates

By email: r.drummond@mmu.ac.uk

21 November 2018

Dear Andrew,

**Re: Confidentiality Letter Agreement between The Manchester Metropolitan University
("Manchester Met") and Andrew Coates (together known as the "Parties")**

On 29 June 2018 the Parties entered into a Confidentiality Letter Agreement ("the
Agreement").

It is now agreed by the mutual consent of the Parties as follows:

1. The provisions in the Agreement shall be varied to delete 'Clause 1 Disclosure'
   and replace with the following wording:

    '1 Disclosure

    *1.1 We, Manchester Metropolitan University (Manchester Met) understand that
    you, Andrew Coates (Recipient) have requested access to Moodle for the specific
    purpose of downloading coursework submitted by students on the Manchester
    Met's courses who have agreed to take part in your PHD research. You have
    obtained from specific students their express consent to access their coursework
    (submitted through Moodle) and your aim is to analyse student textual use in
    coursework. You are a PHD student as well as an Associate Lecturer for the
    Manchester Met and you are currently subject to ongoing obligations of
    confidentiality as a result of such statuses.'*

2. All other clauses in the Agreement shall remain the same and, where specified in
    the Agreement, shall continue to subsist following variation of the Agreement.

3. No deletion, addition or variation to this Letter of Variation shall be valid unless
    agreed in writing and signed by all Parties.

4. This letter of Variation shall be governed by English law and the Parties shall
    submit to the exclusive jurisdiction of the English courts.
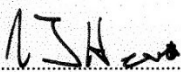
INVESTORS
IN PEOPLE | Champion
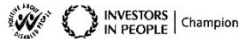
www.mmu.ac.uk
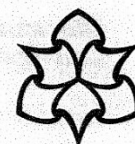
I look forward to receiving the enclosed copy of this legally binding Letter of Variation signed by you by way of acknowledgement and agreement to the terms set out above.

Yours faithfully,

Signed .... Dated ... 27-11-2018 ....

Andrew Hewett
Director of Finance

Acknowledged and agreed by:

Signed .... Dated ... 28/11/2018 ....

Name ... Andrew Coates ....

A6 Extraction Plan

Andrew Coates - Extraction Plan for PG Linguistics Network Student Work

from Moodle

In accordance with the relevant sections of the Confidentiality Letter Agreement, dated 29th June 2018 and signed by myself and Dr Derek Bousfield, the following extraction plan will be followed with regard to accessing the MMU Moodle system and PG Linguistics network students' written coursework:

In accordance with section 1.2.1 - Individual consents from students are stored either as paper documents, or in electronic form, depending on the manner in which students chose to complete them. These are stored in a secure location, or on a secure laptop and encrypted hard drive respectively. A further record of each consent is stored in spreadsheet form so as to keep track of any withdrawals. This is also stored securely. A record of consents will also be supplied to Dr Derek Bousfield, for the purposes of oversight.

In accordance with section 1.2.2 and 1.2.3 – Only work from students who have given written consent to participate in the study (recorded as detailed above) will be accessed.

In accordance with section 1.2.4 – In the event that a student withdraws from the study, accessing of their records will cease immediately.

In accordance with section 1.2.5 – No other information accessible via Moodle will be extracted or used. Only the originally submitted versions of students' work will be downloaded, so scores and feedback will not be included. No areas of the students' records will be accessed save those necessary in order to download coursework submissions as consented to by the students. Participant student ID numbers will only be used to verify consent and identify corresponding submissions. Downloaded submissions will be fully anonymised and securely stored.

Access to the Moodle system will be via a guest staff ID in my name, as obtained by my DOS, Dr Rob Drummond. At present this ID is: 46039218

Project description

Project	The Potential for a Genre and Corpus Analysis of University Students' Academic Writing to Inform Content in the Writing Component of a University Pre-Sessional Course.
Researcher	Andrew Coates – Manchester Metropolitan University

The project is an investigation into the potential for an analysis of students' academic writing to inform course content in the writing element of Manchester Metropolitan University's pre-sessional English course for international students.

The project involves:

1. An analysis comparing academic writing from level 7 students at MMU with published academic writing from comparable disciplines.
2. Interviews and/or questionnaires conducted with students who have completed the pre-sessional course and are now engaged in their programme of study, with the aim of establishing which elements of the pre-sessional writing course they found helpful or unhelpful, and what problems they encounter in the academic writing they are required to do as part of their university courses.

Writing participants – I will access your writing via your anonymous submissions to Moodle. Your anonymity will be protected. I will only use your written text to analyse the use of language. I will not include or consider your grades and feedback in any way. Only myself and my supervisory team will access the data. I will only use your data for this research. The data I collect, and my research in general, will have absolutely no effect on your course outcomes.

Interview participants - Interviews will be audio recorded, then transcribed. Information will be fully anonymised and stored on a secure private laptop and encrypted external drive. Recordings will be destroyed at the conclusion of the project. Your anonymity will be protected, and the opinions expressed will in no way affect your course outcomes. Data will only be accessed by the researcher and supervisory team, and will be used solely for the purposes of this research.

You may withdraw from the project at any time, and you do not have to give a reason. There is no payment for taking part in the project.

For further information, or to withdraw, please email andrew.d.coates@stu.mmu.ac.uk

If you wish to complain about any aspect of the research, you may contact my academic supervisor, Dr Rob Drummond, on r.drummond@mmu.ac.uk

Consent Form

Project:	The potential for a genre and corpus analysis of university students' academic writing to inform content in a university pre-sessional course.
Investigator:	**Andrew David Coates**

Please put a cross in the boxes.

The research project has been fully explained to me.	
I have received a written description of the project.	
I understand that interviews will be recorded, and that recordings will be destroyed at the end of the project.	
I understand that I may request a full transcript of my interview.	
I understand that anonymised verbatim extracts from interviews may be included in publications from this research.	
I understand what I need to do as a volunteer in this project.	
I understand that I will receive no payment for participating in this project.	
I understand that I can withdraw from this project at any time.	
I consent to take part in this project.	

NAME	SIGNATURE

Date: ……………………..

A9 Participant Consent Form (Writing)

![Manchester Metropolitan University logo]

Consent Form

Project:	The potential for a corpus analysis of university students' academic writing to inform content in a university pre-sessional course.
Investigator:	Andrew Coates

Please put a cross in the boxes.

The research project has been fully explained to me.	
I have received a written description of the project.	
I understand that any writing used will be anonymised, and that my grades and feedback will not be used or considered in the study in any way.	
I consent to the use of any written work submitted since my course began and any submitted in the future.	
I understand that anonymised extracts of my writing may be included in publications from this research.	
I understand what I need to do as a volunteer in this project.	
I understand that I will receive no payment for participating in this project.	
I understand that I can withdraw from this project at any time.	
I consent to take part in this project.	

NAME	SIGNATURE

(typing your name constitutes an electronic signature)

Date: ……………………..

MMU Student ID no:………………….…… (used only to identify your submissions)
Course of Study………………………………………
I am a - native English Speaker ☐
I am a – non-native English Speaker ☐   (Please tick as appropriate)

Appendix B Semi-Structured Interviews - Question Schedule

General

a) Which pre-sessional did you do? What are you studying at MMU?

1) Generally, how useful/helpful did you find the pre-sessional course (writing in particular)?

2) How have you found the writing you've had to do so far on your course?

3) Have there been any particular aspects of writing that you have found difficult, or that you feel you are lacking in? Any feedback from lecturers?

4) Do you feel that the pre-sessional course was beneficial in terms of what you need to know for academic writing?

5) Is there anything you would like to have been covered that wasn't?

6) To what extent did the course prepare you in any way for your specific subject area?

Specific Vocab

1) Do you have to use any subject-specific vocabulary when you write? Is this a big part of your writing/learning? How do you typically discover this vocabulary? How do you typically learn it? Are you confident using it?

2) Would you say you have a clear understanding of which words are 'academic' and which aren't? How do you know?

3) Do you feel that vocabulary is something that causes you an issue in your writing? Is this more problematic when you have to write in an academic context?

Specific lexical Bundles

1) Have you had any instruction in terms of words which are generally used together, such as 'last but not least' 'one of the most' 'As a result of', 'it should be noted that'? If so, have you been instructed in which of them may occur commonly in your subject area?

2) Are you aware of which of these constructions might be suitable/unsuitable for academic writing?

3) Do you feel that you are confident using multi-word constructions like this?

4) Do you ever learn words in groups like this, or is it always individual words?

Specific Hedging

1) Have you been taught about 'hedging' (the idea that you should express claims and conclusions to reflect a degree of uncertainty)?

2) Is this something that you think about while writing?

3) Do you feel that you have a good understanding of how to express uncertainty and which methods can be used for this purpose?

4) Do you think this is an important writing skill in your subject area? Why, why not?

Specific Citation

1) Are you confident in your ability to effectively cite previous studies?

2) Do you try to vary the ways in which you cite previous literature?

3) If you vary it, what do you base that decision on?

4) What is your main purpose in citing other work? For example to mention that it exists? To link it to your own work? To evaluate it?

5) When you report other work, how do you refer to it? Neutrally? Positively? Negatively? All of these? What influences this?

6) Do you have any final questions or comments you would like to add?

Appendix C Details of Balancing for Discipline Sub-Corpora

Advertising – balanced for the L2S sub-corpus (49,696 words)

Education – balanced for the L1S sub-corpus (49,870 words)

Fashion – balanced for the L2S sub-corpus (53,246 words)

Forensic Psychology – balanced for the L2S sub-corpus (38,812 words)

Languages – balanced for the L1S sub-corpus (82,230 words)

Physiotherapy – balanced for the L2S sub-corpus (6590 words)

Social Work – balanced for the L2S sub-corpus (19,033 words)

TESOL – balanced for the L2S sub-corpus (262,512 words)