Can IELTS writing scores predict university performance? Comparing the use of lexical bundles in IELTS writing tests and first-year academic writing

Trish Cooper

Wits Language School, University of the Witwatersrand, South Africa
E-mail: Trish.Cooper@wits.ac.za

Abstract
The Cambridge-based International English Language Testing System (IELTS) test is commonly used as one of the criteria in granting students admission to South African universities. It is therefore expected that the language structures elicited by the IELTS tests are representative of the language required in academic writing. Investigating aspects of language, such as the lexicon, should help to assess the validity of the IELTS test as a predictor of university performance. Lexical bundles are groups of words that frequently co-occur and therefore form recognisable clusters that become associated with particular styles of writing. Given that there are marked differences in the use of bundles both within genres and within academic disciplines, university students need to master the lexical bundles particular to their fields in order to display proficiency in the subject in their writing. The study reported on in this paper is based on a corpus of IELTS Task 2 writing tests as well as a corpus of essays, both of which were written by the same first-year psychology students. These corpora were developed to investigate the lexical bundles typical of these two types of academic genres in order to determine the degree of overlap, and therefore assess the validity of Task 2 of the IELTS test as a measure of the style of writing expected from undergraduate students. An analysis of the most frequently occurring 4-word lexical bundles within each corpus was conducted using WordSmith Tools, a computer programme designed for the investigation of aspects of vocabulary within large bodies of texts. The findings show that there are considerable differences in the lexical bundles used as a result of different basic requirements within each essay type. While Task 2 of the IELTS test requires students to present an opinion-based argument, academic essays are based on a number of sources obtained from the relevant literature. Consequently, following from the categorisation of lexical bundles reported in the literature, the lexical bundles in the IELTS test were found to be typical of spoken discourse, whereas those in academic essays are typical of written discourse. A possible implication of this finding is that the current style of the IELTS Task 2 writing test does not serve as a suitable predictor of university performance.

Keywords: lexical bundles, IELTS writing tests, undergraduate academic writing

1.   Introduction

Arguments supporting the concept that lexical items are stored and represented as "multiword sequences" are currently gaining credibility. Consequently, the need to determine the roles and functions of different types of these lexical features is now gaining importance. Following the 1999 study of lexical bundles by Biber, Johansson, Leech, Conrad and Finegan, the results of further research have given credence to the notion that lexical bundles have psychological validity (Jiang and Nekrasova 2007). Given the assumption that a near-native command of lexical bundles is one of the variables that determines proficiency in academic writing (Hyland 2008b:5), this study explores students' use of lexical bundles in academic essays and compares this use to the occurrence of lexical bundles in Task 2 of the International English Language Testing System (IELTS) test.

1.1 Lexical bundles

Lexical bundles are defined as groups of words that commonly co-occur across a range of texts (Biber et al. 1999). As such, they are identified primarily in terms of their frequency of occurrence and distribution criteria. Lexical bundles are also widely recognised as being semantically transparent, thereby distinguishing them from idiomatic expressions, and as having no fixed standard grammatical form in the sense that they occur across grammatical boundaries (Biber, Conrad and Cortes 2004:377). A frequently occurring bundle, for example, is "noun phrase + prepositional phrase[1]" (e.g. "the end of the").

A range of terms has been used in the literature to refer to multi-word expressions, including "multi-word prefabricated sequences/patterns", "formulaic expressions", "fixed expressions", "lexical phrases", "lexical clusters", "lexical chunks" and "lexical bundles" (Biber et al. 2004 and Biber and Barbieri 2007). The key feature that distinguishes lexical bundles from other multi-word sequences is that bundles are defined solely in terms of frequency of occurrence (Biber et al. 2004). Hyland (2008a:44) describes 4-word bundles as being "identified empirically purely on the basis of their frequency rather than their structure". In the original study by Biber et al. (2004), multi-word sequences that occurred at least 10 times in a corpus of one million running words were regarded as lexical bundles. However, this number ranges from 10 to 40 occurrences per million words (Biber et al. 2004; Hyland 2008b).

While frequency is a key identifying feature of bundles, other features must also be considered in conjunction with the rate of occurrence. As a framework for the description of the methodology used in their research, Biber et al. (2004) identified lexical bundles in terms of the following additional criteria: distribution or dispersion threshold, type of bundles (commonly occurring 3-word, 4-word or 5-word bundles), grammatical structure and semantic transparency (as it is possible to determine the meaning of a lexical bundle from the individual words that make up the multi-word sequence).

Another key feature of lexical bundles is that they differ across modes, genres and registers in both form and function (Biber et al. 2004; Cortes 2004; Hyland 2008b). For instance, as illustrated by the lexical bundles found in the aforementioned studies as well as lexical bundles taken from the corpus developed for this study, the majority of bundles in academic writing are

---

[1] The standard abbreviations will henceforth be used (e.g. NP, VP, PP, etc.).

based on NPs (e.g. "the results of the", "the fact that the", "a large number of") and PPs (e.g. "in the case of", "at the end of", "on the basis of"). The distribution patterns of lexical bundles can be used to distinguish between disciplines, as Hyland (2008b:5) elaborates in the following quote:

> These bundles are familiar to writers and readers who regularly participate in a particular discourse, their very 'naturalness' signalling competent participation in a given community. Conversely, the absence of such clusters might reveal the lack of fluency of a novice or newcomer to that community.

Thus, competent, "natural" use of lexical bundles signals or reflects integration into a particular discourse community and marks the user as having a degree of proficiency well beyond that of a novice. From this point of view, Hyland (2008b:5, 8) concludes that the expectations about learners within a particular discourse community include evidence of the ability to use appropriately the lexical bundles specific to that community and that "disciplinary-sensitive" bundles should be identified for classroom teaching practice. This argument provides strong motivation for further studies of both appropriate and inappropriate or non-standard use of lexical bundles in academic writing.

As this study is based on a corpus of students' texts, the review in section 1.2 provides a brief overview of some of the most recognised corpus-based studies of lexical bundles.

1.2 Corpus-based studies of lexical bundles

Various studies have considered the different uses of lexical bundles by students across a range of academic disciplines, with the focus on the distinction between the use thereof by students and professional writers. Cortes (2004) identified lexical bundles in a corpus of history and biology journals. She tagged these as "target bundles" and compared the use of these bundles by academic writers to their use by students. She found, firstly, that students used these target bundles far less frequently than did professional writers. Secondly, she found that, when they were used by students, the functions of these target bundles did not always correspond to those in the journals. Cortes (2004:398) suggests that there are still a number of unanswered questions regarding the use of lexical bundles across different registers, and regarding the extent to which the writing of professional authors differs from that of students across disciplines.

Chen and Baker (2010) compared the occurrence of lexical bundles in published writing with two corpora of student writing – one from first-language (L1) speakers of English and the other from second-language (L2) speakers of this language. The authors found that published academic writing "exhibited the widest range of lexical bundles whereas L2 student writing showed the smallest range" (Chen and Baker 2010:30). More specifically, they found that students overused certain lexical bundles not commonly found in published academic texts while, at the same time, not making use of bundles that might be expected to be familiar as they occur with high frequency in published writing. This contradicts Hyland's (2008a) finding that the majority of lexical bundles identified in his study were drawn from student corpora. Hyland examined the structures and functions of bundles in research articles, doctoral dissertations and masters' theses, and concluded that students are less confident writers and therefore need to rely more on formulaic sequences. However, since research by Chen and Baker (2010), as well

as by Cortes (2004), showed that published writing contains the highest proportion of lexical bundles, this issue clearly requires further investigation.

One of the main aims of Chen and Baker's (2010) study was to investigate lexical bundles in the writing of L2 students of English. The authors note that, while Hyland (2008a, 2008b) considered the writing of L2 students, the participants in his study were at masters' and doctoral levels. Therefore, given their academic proficiency and, correspondingly, their assumed advanced writing skills in academic discourse, these postgraduate students would shed little light on the question of how lexical bundles are used by students with lower levels of English proficiency. The focus of Chen and Baker's (2010) study was therefore more on problems surrounding the use of lexical bundles by L2 students. This concern links directly to Biber and Barbieri's (2007) argument that L2 students cannot succeed within the university context without an adequate command of lexical bundles. These authors recommend that lexical bundles should be explicitly taught in the language classroom and that further studies need to be conducted in order to investigate the acquisition of these features of the lexicon.

In terms of the actual development of a corpus, Biber and Barbieri (2007) provide the formula used to normalise the count of lexical bundles within smaller corpora, while cautioning that distribution must be considered in addition to frequency to avoid idiosyncratic use of language skewing results. Chen and Baker (2010) argue that it is not possible to apply a standardised threshold across all corpora as the conversion of the normalised rate to raw frequencies affects the number of lexical bundles produced. To address this problem, they recommend that both "the raw cut-off frequency and corresponding normalised frequency should be reported" (Chen and Baker 2010:32).

In describing the development of her corpus following the collection of suitable material, Cortes (2004:403) states that journal articles were downloaded in electronic format and "were completely cleaned, that is no titles, headers, footers, captions, scientific formulae, or references were included in the final word count". In order to "clean" their corpus, Chen and Baker (2010:33) removed proper nouns and context-based bundles, such as those containing subject-specific terms. They also merged "overlapping word sequences [which] could inflate the results". These authors identified two types of overlapping sequences (Chen and Baker 2010:33):

> One is 'complete overlap', referring to two 4-word bundles which are actually derived from a single 5-word combination. For example, *it has been suggested* and *has been suggested that* both occur six times, coming from the longer expression *it has been suggested that*. The other type of overlap is 'complete subsumption', referring to a situation where two or more 4-word bundles overlap and the occurrences of one of the bundles subsume those of the other overlapping bundle(s). For example, *as a result of* occurs 17 times, while *a result of the* occurs five times, both of which occur as a subset of the 5-word bundle *as a result of the*.

Chen and Baker (2010) conflated overlapping sequences, representing them with the "fifth" word in the bundle given in brackets, for example, "it has been suggested + (that)" and "as a result of + (the)". Since Hyland's (2008a, 2008b) omission in excluding content-related bundles, proper nouns and "overlapping sequences" may account for the discrepancy in his

findings regarding the use of bundles by students, Chen and Baker's (2010) method of dealing with overlapping sequences is adopted in the present study.

Given that assessing IELTS writing is one of the key aspects of this study, the following section provides an overview of current arguments on the relationship between the IELTS test and academic performance.

1.3 IELTS as a predictor of academic performance

In all the studies conducted to establish the degree of correlation between IELTS results and academic performance, there has been very little general agreement on the value of the IELTS test as a predictor of academic performance (cf. Dooey 1999; Feast 2002; Graham 1987; Woodrow 2006). However, the current viewpoint seems to be that, while low IELTS scores can be used as an indicator of poor language proficiency, too many other variables impact on tertiary education for IELTS scores to be significant (cf. Elder 1993; Hirsh 2007; Woodrow 2006).

The IELTS test is a Cambridge-based assessment of English reading, writing, listening and speaking skills. The academic component of the test is used widely by English-medium universities as an entrance requirement for non-English speaking applicants, with scores ranging in Bands "from zero (did not attempt the test) to nine (expert user)" (Feast 2002:76). The IELTS guidelines recommend a score of Band 6.5 for linguistically less demanding courses, such as mathematics, and a score of Band 7.0 or higher for linguistically demanding courses, such as law (Hirsh 2007:195), in conjunction with the policy that none of the separate skills scores should be lower than Band 6.0. However, "[academic] institutions around the world accept an Overall Band Score of between 6.0 and 7.0 […] as evidence of English language proficiency" (Ciccarelli 2001, in Feast 2002:71). This clearly indicates that a number of international students are accepted into universities and colleges on the basis of lower IELTS scores than is recommended. Given the differences in the minimum scores required for admission, both Graham (1987) and Dooey (1999) suggest that educational institutions should conduct independent studies to establish the link between language proficiency and academic success of their students, and therefore establish assessment thresholds relevant to their respective student bodies, institutional profiles and "the amount of English support available on campus" (Dooey 1999:2). This call for independent research on English proficiency and academic performance serves as further motivation for the present study as the comparison of lexical bundles in the writing component of the IELTS test and academic writing forms the basis of a study on IELTS as a predictor of academic performance.

This study aims to examine lexical bundles that typically occur in student writing within the field of psychology. In addition, the study aims to compare the occurrence of these lexical bundles to their occurrence in Task 2 of the IELTS writing test. To address these aims, the following questions regarding lexical bundles will be considered:

(i)   What are the most frequent 4-word lexical bundles used by students in the field of psychology?
(ii) Are there noticeable differences in the structures and functions of lexical bundles used by students in academic writing as opposed to their use in the IELTS writing tests?

2.   Background to the study

An aspect pivotal to this study is the distinction between the different types of grammatical structures in terms of which lexical bundles are formed and the functions they perform. These categories were initially established by Biber et al. (1999) and are explained in detail in the following sections.

2.1 The structures of lexical bundles

Biber et al.'s (1999) classification system for the structure of lexical bundles, based on the identification of the primary grammatical element within each bundle, has been used in a number of studies in this area (cf. Cortes 2004; Hyland 2008). An adapted version of this framework was developed for the present study in order to account for lexical bundles in the student corpora that did not fall into Biber et al.'s classification system. These adaptations involved the addition of the categories "passive verb + NP", "modal + (passive) VP", "NP", "NP + VP fragment", "NP + *to*-clause fragment" and "comparatives". These categories, as well as those identified by Biber et al. (1999), are illustrated in the examples of 4-word bundles below:

Verb-based:

- VP   ("be able to attend", "be used to explain")
- VP + *to*-clause fragment   ("are more likely to", "can be used to")
- (Passive) VP + PP (fragment)   ("can be defined as", "were included in the")
- Passive verb + NP   ("be given a chance")
- Modal + passive verb   ("should be allowed to", "should be entitled to")
- Modal + VP   ("should be able to", "should have access to")
- Anticipatory *it*   ("it is possible that", "it is important to")
- First-person pronoun
  + dependent clause fragment   ("I agree with the", "I believe that every")

Preposition-based:

- PP   ("on the other hand", "at the same time")
- PP + *of*   ("on the basis of", "at the end of")

Noun-based:

- NP   ("a basic human right", "a significant main
                                                        effect")
- NP + *of*   ("the end of the", "a large number of")
- NP + PP (fragment)   ("a close relationship with", "an important role
                                                        in")
- N(P) + VP fragment   ("this essay will discuss", "research has shown
                                                        that")
- NP + *to*-clause fragment   ("the opportunity to study, "the key to success")
- NP with other post-modification   ("the extent to which", "one of the most")

Lexical bundles with dependent clause fragments:

- *That*-clause fragment   ("that are attributed to", "that there is a")
- WH-clause fragment   ("when it comes to", "which refers to the")

Comparatives: ("as well as the", "is wider than the")

Other: ("because they are not" - conj. + N + VP)

Hyland (2008a:44) supports the argument that types of bundles are clearly linked to register by providing examples of bundles which are characteristic of academic writing as opposed to conversation. He identifies the grammatical features of 70% of bundles found in written academic discourse as:

> preposition + noun phrase fragments (*on the basis of*, *in the case of*), noun phrase + *of*-phrase fragments (*a wide range of*, *one of the most*) as well as anticipatory it fragments (*it is possible to*, *it is clear that*) […] [while] 60% of patterns [in conversation] are personal pronoun + lexical verb phrases (*I don't know what*, *I thought it was*) and auxiliary + active verb (*have a look at*, *do you want a*).
>
> (Hyland 2008a:44)

The lexical bundles in each corpus were analysed in terms of the grammatical structures outlined in this section in order to assess to what degree a comparison can be made between the types of bundles that occurred in the essays written for the IELTS Task 2 test and those that occurred in the undergraduate psychology essays. A similar analysis was carried out in terms of the functions performed by the bundles in each corpus, as discussed in the following section.

2.2 The functions of lexical bundles

Hyland (2008b) retained the grammatical forms identified by Biber et al. (2004) for his own research. However, he developed an alternative set of functions on the grounds that the categories identified by Biber et al. were intended to distinguish between spoken and written modes of discourse, and therefore not all of them are directly applicable to research focused on written registers. Biber et al.'s (2004) study, for example, contains more referential, personal and directive bundles typical of spoken discourse. As the focus is on academic writing, the present study adopts the functions used by Hyland (2008b:13-14), which are presented in Table 1.

Table 1. Types of functions typically performed by lexical bundles in academic texts

No.	Type of function	Description
1	*Research-orientated*	Help writers structure their activities and real-world experiences
1a	- Location	Indicate time/place (e.g. "at the beginning of", "at the same time")
1b	- Procedure	(e.g. "the role of the", "the purpose of the")
1c	- Quantification	(e.g. "a wide range of", "one of the most")
1d	- Description	(e.g. "the size of the", "the structure of the")
1e	- Topic	Related to the field of research (e.g. "of lexical bundles in")

2	*Text-orientated*	Concern the organisation and meaning of the text
2a	- Transition signals	Establish additive or contrastive links between elements (e.g. "in addition to", "in contrast to the", "on the other hand")
2b	- Resultative signals	Indicate inferential or causative relationships between elements (e.g. "as a result of", "it was found that", "these results suggest that")
2c	- Structuring signals	Text-reflexive markers which organise stretches of discourse or direct the reader elsewhere in the text (e.g. "in the next section")
2d	- Framing signals	Situate arguments by specifying limiting conditions (e.g. "in the case of", "on the basis of", "with respect to the")
3	*Participant-orientated*	Focus on the writer or reader of the text
3a	- Stance features	Convey the writer's attitudes and evaluations (e.g. "it is possible that")
3b	- Engagement features	Address readers directly (e.g. "it should be noted", "as can be seen")

The different structures and functions identified in sections 2.1 and 2.2 are used as the basis for the analysis of the lexical bundles in both student corpora. The methodology employed in this analysis is presented in the next section.

3. Methodology

The review of the methodology used in this study is divided into three sections, namely the development of the corpus (section 3.1), the analysis of the lexical bundles (section 3.2) and a profile of the participants (section 3.3).

3.1 Corpus development

The corpus for this study was compiled using Task 2 of the IELTS writing test as well as students' assignments. The IELTS tests were completed by first-year psychology students in tutorials conducted at the start of the first semester. A sample of these written tasks was typed and saved in plain text format for assessment of the lexical features. Any errors made by the students were reproduced exactly in the process.

In order to build a corpus of the students' essays, three sets of assignments – generated by the same students over the course of the year – were scanned to PDF format, converted to a rich-text format, "cleaned" and then converted to plain text for analysis. The "cleaning" process involved the removal of headings and dates as well as references to sources that occurred mid-sentence. This was done to ensure that these elements did not interfere with the identification of strings of 4-word bundles.

The sub-corpus of the students' essays consisted of 208 assignments selected from the submissions of 787 students registered for an introductory course in psychology. This sample represents all the first-year students who had completed the IELTS writing test, a vocabulary test and a demographic survey, and had submitted the three psychology assignments used for the sub-corpus.

Table 2. Number of texts, tokens (running words) and types in each sub-corpus

Description	No. of texts	Words – tokens	Words – types	Lexical bundles – tokens	Lexical bundles – types
May 2011 PSY assignments	200	230 919	6 724 (2.91%)	50 526	4 781 (9.46%)
Aug 2011 PSY assignments	199	290 590	7 696 (2.64%)	32 684	4 009 (12.26%)
Sep 2011 PSY assignments	201	253 914	7 139 (2.81%)	37 070	3 957 (10.67%)
IELTS Task 2 writing	190	55 946	4 001 (7.15%)	6 285	1 907 (30.34%)
TOTALS	790	831 369	–	126 565	–

The computer programme used for identifying lexical bundles was WordSmith Tools (WST) (Scott 2010). In order to identify appropriate bundles, WST was set to identify 4-word combinations of different frequencies, depending on the size of the corpus (ranging from an occurrence of at least twice in the smallest sub-corpus to eight times in the largest sub-corpus), and occurring in a minimum of four texts in the case of the smaller sub-corpora. This calculation of frequency and distribution thresholds was based on the principle of normalisation. Following Chen and Baker's (2010:32) study in which they dealt with similarly small corpora, a minimum occurrence of 25 words per million was normalised by calculating the ratio. Thus, in a text of 100,000 words, the minimum raw frequency would be 2.5 in order to establish the same ratio. As indicated in Table 2, the largest sub-corpus in this study contained 290,590 tokens or running words, while the smallest sub-corpus (that of the IELTS writing tests) contained 55,946 tokens. In terms of the normalisation ratio, the frequency threshold for the largest sub-corpus should be 7.5, while for the smallest it should be 1.4, an option not permitted by WST. Therefore, a decision was made to round off the raw frequencies in all sub-corpora to establish the normalised rate (Chen and Baker 2010) and to set the normalised rate for the IELTS sub-corpus at 2, the lowest option possible. The number of texts in which the lexical bundles occurred was as extensive as possible in relation to the total number of texts in each sub-corpus. The low frequency threshold was therefore counterbalanced by ensuring that the lexical bundles had as wide a distribution as possible, with an average range of 25% across the texts in the corpus. However, the distribution of lexical bundles in the IELTS sub-corpus was fairly low (ranging from 2.63% to 11.05%) as many of the lexical bundles contained keywords from the assignment topic and were therefore disregarded following Chen and Baker's (2010, cf. section 1.2 of this paper) suggestion.

The WST programme was also set to ignore all numbers in the texts, to consider every possible sequence of 4-word combinations without disregarding high-frequency lexical items, such as "the", and to take punctuation into account so that any word sequences spanning a punctuation boundary were not included in the count.

3.2 Analysis of lexical bundles

Each sub-corpus was analysed separately to determine the most frequently occurring lexical bundles in the corpus as a whole. In line with Chen and Baker's (2010) methodology, all lexical

bundles containing words directly related to the essay topic were manually excluded from the count in order to avoid skewing the results in favour of context-specific terms. For example, the topic of the IELTS writing test was "Some people believe that a university education should be available to all students. Others believe that higher education should be available only to good students". To eliminate content words, any lexical bundles that contained keywords from this topic, such as "university", "education" and "students", were not included in the count. Similarly, as the topic of the May psychology assignment was the structure and function of the brain, all lexical bundles containing terms relating to this topic, such as "brain", "hemisphere", "lobes" and "frontal", were disregarded in the count. This was done in order to identify the more generic lexical bundles that are typical of the field of study but not necessarily specific to a particular subject, with the view of arriving at a more valid comparison of different texts.

Furthermore, following Chen and Baker (2010:33), overlapping bundles were merged in order to include the most frequent in the count, with the additional words indicated in brackets. For example, the bundles "should be given the" and "given the opportunity to" were merged to form the single multi-word sequence "should be given the [opportunity to]". In other words, to avoid duplicating the count, the most frequently occurring 4-word string within the 6-word bundle was only counted once in the form of the 4-word bundle "should be given the", with the full string indicated in brackets and with variations on this bundle excluded from the count.

In line with Hyland (2008b), who investigated only the top 50 bundles in depth, the present study focuses on the top 15 lexical bundles across the four sub-corpora (i.e. the IELTS writing test and the three students' assignments). The primary reason for this restricted focus is that this is an exploratory study intended to highlight areas of concern for future research and raise issues that may impact on a more detailed study of these lexical features.

3.3 Participants' profile

The participants in this study were 208 first-year psychology students registered at the University of the Witwatersrand in Johannesburg. Their ages ranged from 17 to 30, although the majority (78.1%) were between 18 and 20 years old, while 16% ranged between 21 and 22 years of age. Most of the students in the sample population were female (87.5%), with only 12.5% of the students being male. English was the mother tongue of 119 (57.2%) of the students, while 89 (42.8%) spoke English as a second or third language.

4.   Results and discussion

Presented below are the results, as well as a discussion, of the analysis of the lexical bundles used by the students in the IELTS writing tests and in academic writing.

Table 3 presents lists of the 15 most frequent 4-word lexical bundles from the different sub-corpora. Less commonly occurring extensions of 4-word bundles are given in square brackets, with the raw frequency and distribution of each bundle given in round brackets. As can be seen from the data in Table 3, none of the lexical bundles listed is common to all sub-corpora; only one of the bundles is common to all three psychology sub-corpora ("as well as the"), while two bundles occur in two of the three student sub-corpora ("as a result of" and "between the ages of"), and only one of the lexical bundles in the IELTS list also occurs in the student assignment sub-corpus ("on the other hand").

Table 3. The 15 most frequent lexical bundles in the IELTS and student assignment sub-corpora

IELTS writing test	May assignments	August assignments	Sept. assignments
should be given the (26; 21)*	[at] the back of the (179; 122)	*between the ages of (79; 52)*	*as a result of (105; 71)*
given the opportunity to (21; 16)	is wider than the (124; 83)	from the age of (69; 42)	*as well as the (66; 49)*
be given a chance (20; 14)	the left side of [the] (102; 73)	*as a result of (63; 43)*	this theory is that (64; 43)
should be allowed to (20; 17)	is located at the (101; 77)	*as well as the (48; 40)*	the way in which (63; 41)
I agree with the (18; 17)	larger on the left (100; 70)	in the face of (46; 40)	*on the other hand (54; 42)*
has the right to (15; 13)	the right side of [the] (91; 66)	*on the other hand (46; 37)*	the world around them (53; 44)
have the ability to (15; 10)	the front of the (81; 75)	who they are and (43; 39)	are learned from the (51; 47)
everyone should be given (14; 10)	is responsible for the (74; 57)	can be applied to (41; 33)	Sense of the world (51; 45)
on the other hand (14; 12)	*as well as the (72; 48)*	at this stage the (38; 28)	to the development of (49; 41)
should be able to (13; 12)	wider than the right (69; 67)	in the case of (35; 25)	according to this theory (48; 36)
to be able to (12; 12)	the functions of the (67; 55)	can be defined as (32; 26)	Make sense of the (48; 44)
to cope with the (12; 12)	different parts of the (60; 50)	is the ability to (32; 31)	*between the ages of (43; 41)*
when it comes to (12; 11)	is divided into two (59; 51)	the next stage of (31; 28)	has been criticised for (43; 32)
will be able to (11; 8)	the side of the (59; 56)	the way in which (31; 26)	does not account for (38; 29)
do not have the (10; 9)	is made up of (46; 34)	is known as the (28; 16)	are expected to be (37; 30)

In his study of variation in lexical bundles across different disciplines, Hyland (2008b:12-13) observed that very few of the bundles identified were common to the fields of applied linguistics, business studies, biology and electrical engineering. He found that only five of the top 50 bundles (10%) occurred across all four disciplines, while 10 of the top 50 bundles (20%) occurred across three disciplines, and more than half of the bundles only occurred in one discipline. Despite the differences in the areas of psychology on which the sub-corpora for this study are based, it was anticipated that a number of lexical bundles would be common across the assignments since the sub-corpora are derived from the same field. However, as was previously mentioned, an analysis of the data in Table 3 shows that only one of the lexical bundles is common to all three psychology sub-corpora ("as well as the"). Part of the reason for this is the very specific nature of the May assignment which focused on the structure and functions of the lobes within the brain, and thus required precise descriptions. The result is a predominance of bundles that differ from those typically used in argumentative essays. This

difference will be illustrated in the discussion of the structures and functions of lexical bundles in the following section.

4.1 Structural types of lexical bundles

The results in Table 4 indicate the breakdown of the 15 most frequent lexical bundles into structural types within each sub-corpus, based on an adaptation of Biber et al.'s (1999) framework (cf. section 2.1). For each different structural type according to which the lexical bundles were assessed, the number of occurrences in the two contrasting sub-corpora (IELTS writing and assignment writing) was counted. The frequency of each structural type is given in terms of the number of occurrences as well as a percentage of the overall number of lexical bundles. As can be seen in line 1 of Table 4, the corpus as a whole contains nine lexical bundles that occur as VPs. These bundles constitute 4.5% of the 200 lexical bundles that occur in the corpus of both IELTS writing and the students' assignments. Of the nine VPs, four occurred in the IELTS test, making up 8% of the total number of lexical bundles in the IELTS sub-corpus, while five occurred in the sub-corpus of the three student assignments, making up 3.3% of this sub-corpus. The percentages serve to provide a common means of comparison across the three sets of results, namely the overall frequency, the IELTS sub-corpus and the assignment sub-corpus. This comparison of lexical bundle types on the basis of frequency as a percentage follows the methodologies used by Hyland (2008a, 2008b) and Chen and Baker (2010) to analyse the occurrence of lexical bundles in different texts.

Table 4. Structural types of the lexical bundles in two different types of student writing

	Structural type	Frequency (Number of occurrences)		No. of uses in IELTS tests		No. of uses in student assignments	
VERB-BASED	VP	9	(4.5%)	4	(8%)	5	(3.3%)
	VP + *to*-clause fragment	21	(10.5%)	9	(18%)	12	(8%)
	(Passive) VP + PP (fragment)	31	(15.5%)	5	(10%)	26	(17.3%)
	(Passive) VP + NP	1	(0.5%)	1	(2%)	0	(0%)
	Modal + passive verb	8	(4%)	7	(14%)	1	(0.6%)
	Modal + VP	4	(2%)	3	(6%)	1	(0.6%)
	Anticipatory *it*	1	(0.5%)	0	(0%)	1	(0.6%)
	1st pron + dependent clause fragment	6	(3%)	6	(12%)	0	(0%)
	TOTALS – verb-based bundles	**81**	**(40.5%)**	**35**	**(70%)**	**46**	**(30.7%)**
PREP.-BASED	PP	22	(11%)	4	(8%)	18	(12%)
	PP + *of*	21	(10.5%)	0	(0%)	21	(14%)
	TOTALS – prep.-based bundles	**43**	**(21.5%)**	**4**	**(8%)**	**39**	**(26%)**
NOUN-BASED	NP	1	(0.5%)	1	(2%)	0	(0%)
	NP + *of*	17	(8.5%)	0	(0%)	17	(11.3%)
	NP + PP (fragment)	4	(2%)	0	(0%)	4	(2.6%)
	N(P) + (passive) VP	16	(8%)	3	(6%)	13	(8.6%)
	NP + *to*-clause fragment	3	(1.5%)	3	(6%)	0	(0%)
	NP with other post-modification	13	(6.5%)	0	(0%)	13	(8.6%)
	TOTALS – noun-based bundles	**54**	**(27%)**	**7**	**(14%)**	**47**	**(31.3%)**

DEPENDENT CLAUSE FRAGMENT	*That*-clause fragment	3	(1.5%)	1	(2%)	2	(1.3%)
	WH-clause fragment	5	(2.5%)	1	(2%)	4	(2.6%)
	TOTALS – dependent clause fragments	**8**	**(4%)**	**2**	**(4%)**	**6**	**(4%)**
	Comparatives	10	(5%)	1	(2%)	9	(6%)
	Others	4	(2%)	1	(2%)	3	(2%)
	TOTAL	200	(100%)	50	(100%)	150	(100%)

As can also be seen from Table 4, there are discrepancies between the types of lexical bundles used in the IELTS writing tests and the academic essays written during the course of the year. The majority of bundles used in the IELTS essays contain VP fragments (70%), redolent of the discursive, argumentative nature of spoken discourse (Biber et al. 2004). This is in strong contrast to the structure of bundles used in the academic essays, where only 30.7% of all bundles used are based on VPs. Given that this is an exploratory study, only margins of 10% or greater between the percentage of lexical bundles occurring in the IELTS texts and the corresponding percentage in the student assignments were considered. Based on this margin, the structures "VP + *to*-clause fragment" (e.g. "will be able to"), "modal + passive verb" (e.g. "should be given the") and "first-person pronoun + dependent clause fragment" (e.g. "I agree with the") in particular occur far more commonly within the IELTS sub-corpus.

Another obvious difference is that the number of preposition- and noun-based lexical bundles that occur within the student assignment sub-corpus contrasts with the proportion of these lexical bundles in the IELTS sub-corpus. While 26% of the lexical bundles within the assignment sub-corpus are preposition-based, only 8% of the bundles in the IELTS sub-corpus are based on PPs. More notably, while 14% of the bundles in the assignment sub-corpus are based on the structure "PP + *of*" (e.g. "as a result of"), none of the bundles in the IELTS sub-corpus are based on this construction. Similarly, while 11.3% of the lexical bundles in the assignment sub-corpus have the structure "NP + *of*" (e.g. "the next stage of"), this construction does not occur within the IELTS sub-corpus. The result is that, in an inverse relationship to the proportion of verb-based bundles to preposition- and noun-based bundles within the IELTS sub-corpus, the occurrence of preposition-based and noun-based bundles is far more dominant in the assignment sub-corpus, together forming 57.3% of the total number of lexical bundles as opposed to the 22% that occur within the IELTS sub-corpus.

4.2 Functional types of lexical bundles

The differences in the functions of lexical bundles used in the students' assignments and in the IELTS writing tests are illustrated in Table 5. As with Table 4, the frequency of each type of function is given in terms of both the number of occurrences and the percentage of the total number of lexical bundles within each sub-corpus. This percentage provides a means of comparing the frequency of the lexical bundles within sub-corpora of different sizes. As with the analysis of structural types, only those lexical bundles with a difference of 10% or greater are included in the discussion.

Table 5. Functional types of the lexical bundles in two different types of student writing

Functions – main categories	Functions – Sub-categories	Frequency (No. of occurrences)		No. of uses in IELTS tests		No. of uses in student assignments	
RESEARCH-ORIENTATED	Location	36	(18%)	2	(4%)	34	(22.6%)
	Procedure	15	(7.5%)	0	(0%)	15	(10%)
	Quantification	13	(6.5%)	0	(0%)	13	(8.6%)
	Description	51	(25.5%)	13	(26%)	38	(25.3%)
	Topic-related	12	(6%)	3	(6%)	9	(6%)
	SUB-TOTALS	**127**	**(63.5%)**	**18**	**(36%)**	**109**	**(72.6%)**
TEXT-ORIENTATED	Transition signals	7	(3.5%)	2	(4%)	5	(3.3%)
	Resultative signals	11	(5.5%)	0	(0%)	11	(7.3%)
	Structuring signals	3	(1.5%)	0	(0%)	3	(2%)
	Framing signals	17	(8.5%)	2	(4%)	15	(10%)
	SUB-TOTALS	**38**	**(19%)**	**4**	**(8%)**	**34**	**(22.6%)**
PARTICIPANT-ORIENTATED	Stance features	31	(15.5%)	27	(54%)	4	(2.6%)
	Engagement features	1	(0.5%)	0	(0%)	1	(0.6%)
	SUB-TOTALS	**32**	**(16%)**	**27**	**(54%)**	**5**	**(3.3%)**
Other		3	(1.5%)	1	(2%)	2	(1.3%)
TOTAL		200	(100%)	50	(100%)	150	(100%)

One of the most apparent differences in the occurrence of functional types is the predominance of research-orientated lexical bundles in the students' assignments, particularly those bundles serving locational and procedural functions. While only 36% of the lexical bundles in the IELTS sub-corpus serve a research-orientated function, 72.6% of the bundles in the assignment sub-corpus serve this function. The primary reason for this difference is that the May assignment required anatomical descriptions of the brain, including the precise position of each lobe, as well as an explanation of where particular functions are situated. These requirements account for the extensive use of lexical bundles that perform the location function. The students were also required to describe the functions of the brain and therefore drew on lexical bundles that perform the procedural function, such as "is responsible for the" and "the purpose of the".

Another clear difference is the frequency of occurrence of text-orientated lexical bundles in each sub-corpus. Although none of these functions have a difference of greater than 10% when considered separately, the overall sub-total for these types of bundles shows a difference of 14.6%. The reason for this difference is clearly the particular prevalence of framing signals, such as "in the case of", and resultative signals, such as "as a result of", in the students' assignments; however, these types of functions were not commonly used in the IELTS tests, with the framing signals accounting for only 4% of the total number of lexical bundles and the resultative bundles not occurring at all.

Finally, the most obvious difference in the occurrence of functional types is the high frequency of lexical bundles that perform a stance function in the IELTS sub-corpus. As this function is used to convey the writer's attitudes, it is to be expected that an essay requiring the writer to express his or her opinions will contain a high proportion of stance-based lexical bundles, such as "I agree with the" and "everyone should be given". This is confirmed by the fact that 54% of the lexical bundles in the IELTS sub-corpus perform a stance function. The assignment topics, on the other hand, required students to provide descriptions, explanations and arguments

on the basis of established theory and accepted opinion. The student was not encouraged to express his or her opinions freely or without referring to recognised sources, hence only 2.6% of the lexical bundles in the assignment sub-corpus perform the stance function.

As this study is based on a fairly small corpus, the findings should be regarded with a degree of caution. Nevertheless, there is a reasonably considerable difference in the use of lexical bundles that indicate location in the student assignment sub-corpus, although this could be explained by the nature of the May assignment which required students to provide more direct references to place and position, as discussed in section 3.2.

As with the structural types, there is a notable difference between the functions used in IELTS writing and those employed in the students' academic writing. In keeping with the finding that the majority of structural types in the IELTS sub-corpus are based on fragments of VPs rather than on NPs, the majority (54%) of these lexical bundles perform a stance function, thus "conveying the writer's attitudes and evaluations" (Hyland 2008b:14). This is to be anticipated given that the nature of the essay topic required students to present their own opinions without reference to outside sources. The assessment of these bundles as VP-based provides further evidence for the argument by Biber et al. (2004) that there is a strong relationship between the structures and functions of lexical bundles. The 27 bundles that perform a stance function in the IELTS essays are all fragments of VPs, with lexical chunks, such as "should be + V" and "I agree with" used to present the students' opinions. Only one of the 15 lexical bundles from the IELTS list in Table 3 performed a function more commonly associated with academic writing, that is, transition ("on the other hand").

One of the research questions presented at the start of this paper addressed the issue of whether there are noticeable differences in the types of lexical bundles used by students in academic writing as opposed to in the IELTS writing tests. The analysis of lexical bundles revealed a considerable difference between the types used in each sub-corpus. While the lexical bundles in the students' assignments tended to be based on NPs and PPs, and are therefore characteristic of academic writing, the bundles in the IELTS essays were based on VP fragments, such as those more typically found in spoken discourse. In line with this type of structure, these bundles performed a stance function as the students presented their own opinions on the topic. One of the primary distinctions between the IELTS essay and the academic assignment is that, while the assignment required students to draw arguments and opinions from a range of sources and then represent these from their own perspectives, the IELTS essay simply required students to argue a point based entirely on their own experientially-based opinions without any reference to acknowledged sources. The variations in the types of lexical bundles in these two genres clearly reflect this difference and, consequently, questions are raised regarding the validity of the IELTS writing test as a predictor of academic performance.

5.   Conclusion

Given that research into lexical bundles is in its infancy, there are a number of unanswered questions surrounding this aspect of vocabulary acquisition. Biber et al. (2004:372) note that "[t]he overall importance of multi-word units in discourse can be fully understood only by undertaking empirical research studies from different perspectives".

As was previously discussed, the present study has demonstrated that there are notable differences in the structures and functions of lexical bundles used in Task 2 of the IELTS writing test and in students' undergraduate assignments. The nature of these differences suggests that the style of writing expected in the IELTS test does not correspond to that expected in student assignments. This leads to the conclusion that the opinion-based essay required for Task 2 of the IELTS writing test does not serve as a predictor of students' abilities to produce academic writing.

This study has contributed to the research on lexical bundles and has added to the growing body of knowledge on how bundles are used in academic disciplines. However, further research is required on the types of lexical bundles that occur in fields other than psychology as this would shed light on the question as to whether certain lexical bundles are generic to all academic fields while others are subject-specific. In addition to this, further research should also be conducted on the differences between the lexical bundles used by students and professional writers. This would help students gain a command of the register by identifying ways in which the gap between student writing and professional writing could be reduced. There is a general consensus that L2 learners struggle with the acquisition of lexical phrases in particular and that learning to use the lexical phrases specific to a discipline contributes to a sense of communicative competence within that field of study. The analysis of such fixed phrases within each discipline therefore helps the learner acquire the expected rhetorical tools.

Another key research area involves the validity of the IELTS test as a predictor of academic performance. Although many South African universities use the IELTS test as a gate-keeping mechanism, the discrepancies in writing styles between Task 2 of the IELTS writing test and an academic essay suggest that the demands of this type of IELTS question do not necessarily reflect the requirements of argumentative writing. It may be advisable for these universities to identify or develop an alternative writing test that more accurately assesses a candidate's ability to produce writing in line with the expectations of undergraduate assignments.

References

Biber, D. and F. Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26(3): 263-286.

Biber, D., S. Conrad and V. Cortes. 2004. "If you look at …": Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3): 371-405.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *The Longman grammar of spoken and written English.* London: Longman.

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

Cortes, V. 2004. Lexical bundles in published and disciplinary student writing: Examples from history and biology. *English for Specific Purposes* 23(4): 397-423.

Dooey, P. 1999. An investigation into the predictive validity of the IELTS test as an indicator of future academic success. *Teaching and Learning Forum.* Available online: http://lsn.curtin.edu.au/tlf/tlf1999/dooey.html (Accessed 14 September 2012).

Elder, C. 1993. Language proficiency as a predictor of performance in teacher education. *Melbourne Papers in Language Testing* 2(1): 68-87.

Feast, V. 2002. The impact of IELTS scores on performance at university. *International Education Journal* 3(4): 70-85.

Graham, J.G. 1987. English language proficiency and the prediction of academic success. *TESOL Quarterly* 21(3): 505-521.

Hirsh, D. 2007. English language, academic support and academic outcomes: A discussion paper. *University of Sydney Papers in TESOL* 2(2): 193-211.

Hyland, K. 2008a. Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18(1): 41-62.

Hyland, K. 2008b. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27(1): 4-21.

Jiang, N. and T.M. Nekrasova. 2007. The processing of formulaic sequences by second language speakers. *The Modern Language Journal* 91(3): 433-445.

Scott, M. 2010. *Wordsmith Tools 5.* Oxford: Oxford University Press.

Woodrow, L. 2006. Academic success of international postgraduate education students and the role of English proficiency. *University of Sydney Papers in TESOL* 1: 51-70.