



UNIVERSITY OF LEEDS

This is a repository copy of *Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/150663/>

Version: Accepted Version

Article:

Coxhead, A, Yen Dang, TN orcid.org/0000-0002-3189-7776 and Mukai, S (2017) Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, 30. C. pp. 66-78. ISSN 1475-1585

<https://doi.org/10.1016/j.jeap.2017.11.001>

© 2017, Elsevier Ltd. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Coxhead, A., Dang, T. N. Y., & Mukai, S (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, 30, 66–78.

Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks

Small-group academic interactions, such as tutorials and laboratory sessions, do not often feature in vocabulary or English for Academic Purposes (EAP) research. Yet across many disciplines in tertiary education, students are required to attend and participate in these speaking and listening events. EAP students need to be prepared for them, and their textbooks should help them prepare also. The present study included interview data from lecturers and students which identified specific issues with speaking in small groups in university; corpus-based analysis of tutorials and laboratory sessions which showed large amounts of high frequency vocabulary; and an EAP/ESP textbook analysis showed little focus on vocabulary in tutorials and nothing on laboratories. A total of 176 phrases were suggested in three textbooks as being useful for expressions for speaking in small groups. An analysis of these phrases in the corpus showed that three quarters of these phrases did not appear in the laboratory corpus and two thirds did not appear in the tutorial corpus. Finally, lexical patterns from the corpus were identified and categorised in the same way as existing lists of multiword units in spoken academic English. Implications for pedagogy and materials design are followed by suggestions for future research.

1. Introduction

English for Academic Purposes (EAP) research has focused predominantly on writing for the university (Skyrme, 2010), with some attention in terms of academic speaking focused mostly on lectures (Feak, 2013), particularly with regard to discourse in lectures (Jordan, 1997) and student problems with understanding content in lectures (Basturkmen, 2016; Crawford Camiciottoli, 2010). Student presentations have been the focus of some research (e.g., Hincks, 2003), and are an area of concern for second language speakers of English (Kim, 2006). Lectures and student presentations are mostly one-way, in that one speaker tends to hold the floor. Basturkmen (2016) notes that there is limited research into academic speaking which involves interaction between speakers, such as between lecturers and students or students and students. This form of ‘dialogic

interaction' is important for developing understanding of disciplinary content and ability to express complex ideas (Basturkmen, 2016). The lexical nature of small-group academic interactions, such as tutorials and laboratory sessions, does not often feature in vocabulary or English for Academic Purposes (EAP) research. Yet across many disciplines in tertiary education, students are required to attend and participate in these speaking and listening events. Labs and tutorials are opportunities for students to deepen their understanding of the concepts, theories, and reading issues in the lectures. EAP students need to be prepared for them, and EAP classes and textbooks should support this preparation. However, EAP courses tend to focus on study skills and EAP textbooks often use a functions-based approach for developing EAP speaking skills (Basturkmen, 2016), in the form of functional expressions such as ways to agree or disagree with someone (e.g. I totally agree) or give opinions (e.g. in my experience/opinion). Do these approaches actually support learners in understanding the vocabulary that they will encounter and have to use in small group interactions at university?

Researchers have also used functions when reporting on the use of lexical bundles (4-word multi-word units) in academic texts. Biber, Conrad, and Cortes (2004), Simpson-Vlach and Ellis (2010) and Wood and Appel (2014) have drawn on stance expressions, discourse organisers, referential expressions, and special conversational functions as a way to identify, categorise and organize lexical bundles in academic spoken and written texts. Do these categorisations feature in spoken tutorial and laboratory corpora?

The purpose of this article is to focus on vocabulary in small group interactions at university through interviews with lecturers and students, a corpus-based analysis of single words and multi-word units in tutorials and laboratories, and an analysis of EAP textbooks

1.1. Speaking in small groups in the university context

Tutorials and laboratories are typically small group learning opportunities which run alongside lectures, the main spoken academic event at undergraduate level. At postgraduate level, interactive seminars are becoming the main spoken academic event (Basturkmen, 2016; Marlina, 2009). Basturkmen (2016) outlines why interaction is important in academic speaking: it provides confirmation of student comprehension of key ideas, supports the development and confirmation of content knowledge for students, helps with relationship development with others in the class (staff and students), and builds familiarity with disciplinary norms.

While Marlina (2009) reported social and educational benefits of tutorial participation in an Australian university by second language speakers of English, Jones (1999) noted that these speaking environments can cause strain for non-native speakers. In a longitudinal study of Chinese learners in a university setting in New Zealand, Skyrme (2010, p. 212) found that “within the university, speaking was the skill least required and most feared by the students, as it subjected them to public evaluation”.

According to Csomay (2006), one distinctive feature of university classroom teaching in general is on-line informational elaboration. This feature characterizes the situation where speakers share the same contexts with listeners, and are under the pressure of transferring the information (informational) in real time production circumstance (on-line). In other words, in small group speaking such as labs and tutorials, students need to process information-dense messages and then give response in a limited period of time. Such pressure means that fluency is the key for these students to successfully engage in small group speaking.

Tutorials are more common in soft sciences (e.g., business, art) (Neumann, 2001) and can be highly interactive environments (Hunter and Coxhead (2007)). Learners might be required to attend a lecture and also read in advance of a tutorial to be prepared for discussion. Tutorials might reflect some flavour of the local context, such as current events, cultural and historical references, and jokes and stories in lectures (Reinhart, 2002), as well as the use of colloquialisms and local lexis (Coxhead, Hunter, Pierard and Cooke (2008)). These elements of academic speaking can provide challenges for even the most competent, confident and prepared speaker of English as a second or foreign language, particularly if they have spent very little time in a particular context or do not have strong networks of support.

Laboratories are more common in hard sciences (e.g., engineering, biology) (Neumann, 2001). The majority of research on the linguistic features in laboratories has focused on analysis of written texts, such as genre analyses of student laboratory reports (see, for example, Kelly-Laubscher, Muna, & van der Merwe, 2017; Parkinson, 2017). To the best of our knowledge, only one study (Tapper, 1994) has focused on the spoken aspect of laboratories. The focus of that study was to examine the use of directives (particularly pronouns) in laboratory sessions by an international teaching assistant in the Australian context. Laboratories appear to be a neglected area of research in EAP.

1.2. Vocabulary in spoken academic contexts

University students need to comprehend both academic spoken and written English: vocabulary knowledge and comprehension are significantly related (e.g., Schmitt, Cobb, Horst, & Schmitt, 2015; van Zeeland & Schmitt, 2013). Corpusbased studies have contributed a great deal to developing our understandings of aspects of academic lexis (both single word and multi-word units). Numerous studies have focused on written discourse (e.g., Ackermann & Chen, 2013; Coxhead (2000); Gardner & Davies, 2014; Lei & Liu, 2016; Watson-Todd, 2017; Wood & Appel, 2014; Yang, 2015) while studies investigating the vocabulary in academic spoken English are very limited in number. The academic spoken English research appears follows three main trends. The first trend is to investigate the vocabulary loads of different kinds of academic spoken English. Dang and Webb (2014) showed that a vocabulary size of 4000 and 8000 word families plus proper nouns and marginal words (such as *um* and *ah*) is needed to reach 95% and 98% coverage of lecturers and seminars in English, respectively. Webb and Paribakht (2015) found similar coverage levels in a study of listening passages (interviews, announcement, dialogues, and short lectures) from the CanTEST, a university admission test in Canada: 4000 word families to reach 95% coverage of academic spoken text and 10,000 word families to reach 98%. The second trend is to determine the coverage of Coxhead (2000) Academic Word List (AWL), developed from a written corpus, in academic spoken discourse. These studies reveal that the AWL provided 2.4% coverage of student presentations (Hincks, 2003), 4.9% coverage of lectures (Thompson, 2006), 4.41% of lecture and seminars (Dang and Webb (2014)), and 6.48% coverage of university admission listening comprehension tests (Paribakht & Webb, 2016). These coverage figures are much lower than the coverage of the AWL in academic written English (10%), which suggests that vocabulary in academic speech is distinctive from that in academic writing. McCarthy and Carter (1997) separating written and spoken corpora for analysis. While these corpus-based studies provide frequency-based information on the lexical items which are used in academic speech in English, but as McCarthy and Carter (1997, p. 30) state, "... computers are less useful when it comes to understanding the way vocabulary is used as a communicative resource by individual speakers in individual situations". Sinclair and Renouf (1988) make the point that data on frequency and usage of vocabulary can be used together to good effect. The third trend is to develop word lists which consist of the words that can help to enhance students' comprehension of academic spoken English. Dang, Coxhead and Webb (2017) Academic Spoken Word List (ASWL) consists of 1741 word families

selected from a spoken academic corpus of just over 13 million running words. This corpus was organized into four equally-sized subcorpora (hard-pure, soft-pure, hard-applied and soft-applied) based on a classification of higher education disciplines from Becher (1989). The ASWL covered 90.13% of the tokens in that corpus and may allow learners to reach 92% e96% coverage of academic spoken English depending on their proficiency levels.

The ASWL is a list of single words. McCarthy and Carter (1997) point out that single words in lists may also appear in multiword units in spoken interaction. Research into multi-word units in spoken academic English by Simpson-Vlach and Ellis (2010) resulted in the development of a spoken Academic Formulas List, which consists of 979 items. This list was developed from a 2.1 million word spoken corpus. This corpus had five unequally-sized sub-corpora: Humanities and Arts (559,912 running words), Social Sciences (710,007 running words), Physical Sciences (363,203 running words), and Nondepartmental/other (159,592 running words).

The linguistic features of academic spoken English vary according to speech events (Biber, 2006; Schmitt et al., 2015), but research has not distinguished between them (Dang et al. (2017); Simpson-Vlach & Ellis, 2010). No research so far has focused on small group speaking such as tutorials and laboratories.

1.3. Research on how EAP textbooks make use of information from corpus-based studies

EAP textbooks play an important part in how language is taught and learned (see Bondi, 2016). It is important to examine how closely EAP textbooks correspond to the vocabulary that learners are likely to encounter and use often in their academic study. Four studies (Chen, 2010; Miller, 2011; Wood & Appel, 2014; Wood, 2010) have investigated this issue. Wood (2010) identified lexical clusters in EAP textbooks, focusing mostly on the language of classroom instruction rather than academic lexical clusters. Miller (2011) examined the coverage of the AWL in ESL reading textbooks and university textbooks and found that the ASWL provided lower coverage (4.78%) in ESL textbooks than in the university textbooks (8.40%). He then raised a concern that textbooks might be "... providing students neither the exposure to the range of academic vocabulary nor the number of encounters with academic vocabulary that they may need to develop successful comprehension of university textbooks" (Miller, 2011, p. 39). Chen (2010) compared lexical bundles in ESP textbooks and university introductory textbooks, and suggested that EAP textbooks failed to reflect the lexical bundles in the university textbooks. Wood and Appel (2014) found that the majority of the multiword constructions in university

textbooks did not occur in the EAP textbooks, and if they did, they were given no pedagogical treatments. All of these studies focused on written discourse. Sinclair and Renouf (1988) report that textbooks tend not to capture pragmatic and discoursal roles of lexis in interactive texts, which is why further investigation into how EAP textbooks prepare students for this task is warranted.

1.4. Research questions

The present study draws together the information from different resources related to EAP learning and teaching (teachers, learners, corpora, and EAP/ESP textbooks). There are five research questions in this study:

1. What do university lecturers and students have to say about speaking in small groups at university?
2. What are the vocabulary loads of labs and tutorials?
3. Do EAP and ESP textbooks focus on vocabulary in labs and tutorials? If so, what aspects of vocabulary do they focus on?

What patterns of vocabulary do they focus on and why?

4. How often do these textbook patterns appear in a spoken corpus of university labs and tutorials?
5. What other lexical patterns appear in a spoken corpus of labs and tutorials?

2. Methodology

2.1. Interview data

Semi-structured interviews were carried out with 24 postgraduate students (nine local and 15 international) across Applied Linguistics, Engineering and the Business School in a New Zealand university. A further 17 lecturers from the same three disciplines also took part in interviews. Table 1 has the details of the participants. Ethics approval for the interviews was granted by the university Human Ethics Committee and all participant names were changed. The interviews focused on interaction and talk in postgraduate classes (Appendix 1). The interviews took approximately one hour each and were recorded. Elements of interest in this article, such as language use and vocabulary in small group interactions such as tutorials, arose through the interviews. Corpora of labs and tutorials. Two corpora were developed for the present study (Table 2). The lab corpus represents 10 subject areas, and contains

137,399 running words from the Michigan Corpus of Academic Spoken English, Limerick-Belfast Corpus of Academic Spoken English, and Newcastle Corpus of Academic Spoken English. The tutorials corpus contains 380,078 running words from nine subject areas from the Limerick-Belfast Corpus of Academic Spoken English and Hong Kong Corpus of Spoken English.

2.2. EAP and ESP textbooks

Fifteen series of EAP speaking and listening course books from various publishers and one series of ESP textbooks were examined (Appendix 2). These textbooks were available to EAP learners at the Language Learning Centre (LLC) at Victoria University of Wellington, based on the recommendation of the EAP teachers, students, and the ratings on the amazon.com website.

2.3. Data analysis

2.3.1. Interview data

The interview data was analysed using a staged system of coding (Corbin & Strauss, 2008), using Grounded Theory (Glaser & Strauss, 1967); a process which was not always linear. The third researcher coded the interview data into broad categories: lecturers' choices and motivations for interactive speaking in class and factors affecting students' participation in interaction in class (see Mukai, 2017). Coding was checked by the first author and also cross-checked with participants to ensure that interpretations of meaning were accurate.

2.3.2. Vocabulary loads of labs and tutorials

Heatley, Nation, and Coxhead's (2002) RANGE program was used with Nation's (2012) 25 BNC/COCA lists. The BNC/COCA lists were developed based on the frequency and range of occurrences of the word families in the British Nation Corpus, the Corpus of Contemporary American English (Davies, 2008), and the Wellington Corpus of Spoken New-Zealand English (Holmes, Vine, & Johnson, 1998). These lists are ranked according to frequency and range. Each level of the BNC/COCA lists consists of 1000 word families, with the 1st 1000 containing the most frequent words. The word families in the BNC/COCA lists were based on Level 6 word family from Bauer and Nation's (1993) taxonomy. For example, the word family for clarify consists of a base word (clarify), its inflections (clarifies, clarifying, clarified) and derivations made up of affixes up to Level 6 in Bauer and Nation's (1993) taxonomy (e.g., clarification, clarity). The Range program has supplementary lists of proper nouns, marginal words,

compound words, and abbreviations. During the analysis, the following items were not found in the lists: 231 proper nouns (e.g., Barack, Branton), nine marginal words (ooooh, haha), 16 compound words (pumpkinseed, checkbox), and 11 abbreviations (e.g., ZDP, USB). These items were added to the relevant lists.

2.3.3. Textbook analysis

The textbooks were examined to see if there are any textbooks focusing on vocabulary in labs and tutorials, and if so, the lexical patterns introduced in these textbooks and their functions were counted.

2.3.4. Corpus analysis

Antconc (Anthony, n.d.) was used to identify the relative frequency of each pattern per million, calculated by dividing the raw frequency of the pattern in the corpus by the size of the corpus and multiplying by 1000. For example, because I agree with occurred three times in the lab corpus (137,399 running words), and its relative frequency in the labs corpus was 28.83. Three lists of 4-word word strings were generated from the corpus of labs, tutorials, and labs plus tutorials. A four-word sequence is the most common way to identify lexical bundles (Simpson-Vlach & Ellis, 2010; Wood & Appel, 2014). Following Wood and Appel (2014), we chose a frequency cut-off point of 25 occurrences per million words because it is between the range of frequency cut-off point used in previous studies into lexical bundles (20-40 occurrence per million words). To deal with the overlap among these clusters, we broke each 4-word cluster (e.g., the rest of it) into two constituent 3-word clusters (e.g., the rest of, rest of it). The frequency of the two 3-word clusters in the corpora were identified and compared. If the frequency of one 3-word cluster was at least double the frequency of the other, the more frequent cluster was classified as the root structure and the 4th words was considered as a word that commonly occurred with that structure and was put in a bracket. For example, in the tutorial corpus, the rest of occurred more than three times (freq $\frac{1}{4}$ 48) than rest of it (freq $\frac{1}{4}$ 15). Therefore, the final structure is the rest of (it). If any two 4-word clusters contained identical 3-word clusters, the 3-word cluster shared among these 4-word clusters were classified as the root structure and the remaining word in each 4-word cluster was placed in a bracket. For instance, at the end of and the end of were represented as (at) the end (of).

3. Results

3.1. Research question one: What do university lecturers and students have to say about speaking in small groups at university?

Data from the interview shows that small group discussion is encouraged by lecturers, as Ian, a lecturer in Applied Linguistics, says.

I deliberately yeah I do encourage people Sem is very competent. I like his perspective so definitely I like to hear what he has to say, but he doesn't always speak up Amy's got great ideas but is much shyer so I got something, different reasons for asking you, because I want to, her to have an opportunity to have her voice.

Tim, a lecturer in Engineering, believed that interaction benefitted concentration levels. He remarked,

If I go 10-15 min in the class without student interaction I think that's too long ... Not a really good student can do 20 min with solid concentration and absorb new information. So what I tend to do is ... after a 10-15 min have something that allows them to re-absorb all the information I've given them and a discussion or an activity is a good way to do it. (Tim, Lecturer, Engineering).

Another benefit of interaction emerging from the interviews with lecturers is to consolidate and deepen students' content knowledge. Here is Tim from Engineering again, who tells his students,

Go think about it yourself first. Discuss it amongst yourselves, then go turn to a person that sits next to you, and verbalize what you are thinking. Because if you can't explain what you are thinking then you don't really understand it.

Brenda, an Applied Linguist, echoed Tim's point. She said,

just talking to somebody and advancing your idea of it 'cause that actually makes you take sort of steps far for-forward that you might not have done if you're just thinking about it.

Hannah, another lecturer in Applied Linguistics, expressed this idea as follows,

I think it's very important to have these group or pair activities so that they can learn to verbally express their thoughts in the way that gets what they want to say and what they are meaning across to another person, and the more they do that, the better at transferring into their writing skill as well, because they start paying attention to it and yeah what information needs go there.

The interviews with the students revealed that they experienced difficulties with processing content in small group discussions. Fluency was raised as a particular problem for international students. For example, Isaac, from China who was studying Finance summarized the problem this way,

Sometimes ... when I think about that how to translate the answer into English, someone else will give the answer.

Yeah and I will miss the chance. And on the other hand, if I just speak out, if I didn't prepare it and just speak out my answer in English, sometimes it would it would be puzzled or confusing, so I need to explain again and again ... you know we need to spend more time to construct our answer because we need to think about the answer in my mother language and we need to translate it in English.

Another problem came from the interactive nature of small group discussions, according to Jasmine, a Taiwanese international student in Applied Linguistics, who said,

They [native speakers] will connect sounds when they speak fast. They connect words, really squeeze words together.

Isaac found it difficult to follow the talk . He said,

sometimes if you are a little bit frustrated in communicate with Kiwi students, ... sometimes they speak too fast and they never open their mouth, so ... a lot of time we just sit down and chat with each other. I cannot even catch up with the speed.

These difficulties were not just one way, as Josh from Papua New Guinean who was studying Engineering noted,

Maybe sometimes they, locals, they don't understand the way we speak. When I speak to them, they couldn't quite understand me. They ask me to explain what I really mean. Yeah, that's a thing I always, they don't seem to understand the way you, maybe because of the accent.

Cultural differences in approaches to speaking were a problem for Grace, from Hong Kong, who was studying Applied Linguistics. In her experience, New Zealand students tended to take the lead in conversation in class, and Asian students tended to be more reserved. Ruby from Taiwan noted,

When it is native speaker and non-native speaker, I think mostly that native speaker will try to keep checking if the non-native speaker follow up his or her thought, and does he, the non-native speaker agree ... Usually native speaker lead the conversation, so kind of more towards native speaker way, ' cause we just tend to be like “ yeah yeah yeah, agree, yeah, that's right” .

The fourth reason is demands of frequent turn-taking. Rachel, a local student in Applied Linguistics, identified turn taking as a problem in academic speaking from her experience. She stated,

I think, in general, in the classes I've been in this year, uhm, I think some native speakers will have a longer turn, maybe, speak a little bit more longer, and try to, uhm, maybe, I

think, yeah, possibly some of them are more confident to, sort of, express themselves and, uhm, yeah, it seems like they might be talking about, length-, more lengthy.

Vocabulary was not seen as an immediate cause of concern in academic small group interactions for students. One exception is Bin, a Vietnamese Engineering student, who said,

... the pronunciation and using like a collocation, it's not correct So that's the difficulty we have to improve, in terms of using words and using collocation.

While the interview data from lecturers and students provides plenty of food for thought on motivations for and barriers of interactive speaking in university settings, little evidence comes to light that suggests vocabulary is a major cause of concern. Let's turn to the corpus analysis to find out more about the nature of lexis in laboratories and tutorials, to see what light this data might be able to shed.

Research question two: What are the vocabulary loads of labs and tutorials?

Table 3 presents the lexical coverage provided by each level of Nation's (2012) BNC/COCA lists in labs and tutorials. The 1st 1000 word families accounted for 85%-87% of the words, followed by the 2nd 1000 word families (4%-5%), and the 3rd 1000 (2% -3%). From the 5th 1000 word families, the coverage dropped to under 1% for each frequency level. The percentage of proper nouns and marginal words in these speech events is higher than the coverage provided by the 4th 1000 word families.

According to Nation and Webb (2011), the learning burden of these words is less than other words. If we assume that learners can recognize proper nouns and marginal words, we can add these figures to the coverage provided by word families at each level, (see, Dang and Webb (2014); Nation, 2006; Webb & Rodgers, 2009a, 2009b).

Table 3 presents the cumulative coverage provided by the BNC/COCA lists plus proper nouns and marginal words in tutorials and labs. It shows that labs are more demanding than tutorials in terms of lexical coverage. If proper nouns and marginal words are counted, a vocabulary size of 3000 word families is needed to reach 95% coverage of labs and a vocabulary size of 7000

word families is necessary for 98% coverage. In contrast, knowledge of the most frequent 30 0 0 word families and the most frequent 70 0 0 word families plus proper nouns and marginal words reach nearly 98% and nearly 10 0% coverage of tutorials. The most frequent 20 0 0 word families provided nearly 95% coverage of tutorials and the most frequent 10 0 0 BNC/COCA word families covered about 86% of the words in labs and tutorials.

3.2. Research question three: Do EAP and ESP textbooks focus on vocabulary in labs and tutorials?

The textbook analysis revealed that none focused on vocabulary in labs. This is not since there were no specifically sciencebased textbooks in this study. Only three series in the set of texts analysed, Oxford EAP, Communicating on Campus, and EAP Now, had sections focusing on vocabulary in tutorials. These three series introduced 176 phrases briefly in the form of useful expressions for speaking activities, such as checking for comprehension/asking for clarification/confirmation/elaboration (e.g., Could you please say that a little more slowly/loudly, Could/can you clarify); offering a fact or example (e.g., let me give you an example, I'd like to add); and paraphrasing (e.g., in other words, what she meant was) (see Appendix 3 for all of the categories and functions). But do these expressions in the textbooks appear in the labs and tutorials corpus?

3.3. Research question 4: How often do these textbook patterns appear in a spoken corpus of university labs and tutorials?

Table 4 presents the relative frequency per million of the 176 patterns in (a) the lab corpus, (2) the tutorial corpus, and (3) the corpus made of both labs and tutorial. For example, the second row of the table shows that 5 out of the 176 patterns (2.84%) had a relative frequency from 1 to 9 per million in the lab corpus. The number of items with relative frequency ranging from 1 to 9 per million in other corpora was 22 out of 176 (12.5%) (tutorial corpus) and 27 out of 176 (15.34%) (both tutorials and labs). What is apparent from Table 4 is that most of the 176 patterns did not appear in the three corpora. The first row shows that 73.86% of these patterns (130 out of 176) did not appear in the lab corpus. Note that 64.77% of the 176 patterns (114 out of 176) did not occur in the tutorial corpus and 63.07% (111 out of 176) did not occur in the corpus made up of both labs and tutorials.

3.4. Research question 5: What other lexical patterns appear in a spoken corpus of labs and

tutorials?

Tables 5-7 present some examples of the sequences appearing in the corpora of labs (Table 5), tutorials (Table 6), and labs plus tutorials (Table 7), respectively. A fuller list can be found as supplementary materials on the journal's website. These sequences are categorized into four functional groups based on Biber, Conrad, and Cortes's (2004) functional classification of lexical: stance expressions, discourse organizers, referential expressions, and special conversational functions. Stance expressions (e.g., I want to) convey attitudes (e.g., desire, obligation, ability) or degree of certainty toward subsequent proposition.

Discourse organizers (e.g., if you look at) indicate the relationships between prior and coming discourse (e.g., introducing and elaborating topics). Referential expressions (e.g., the rest of the) refer directly to physical or abstract entities, or to the textual context itself. Special conversational functions (e.g., what's going on) express politeness, simple inquiry, or reporting. In Tables 5-7, for each 4-word sequence, the root structure is highlighted in bold italics while the remaining word in the sequence is placed in a bracket. Condensing overlapping structures in this way makes it possible to see the base structure of the sequence and the words that are most likely to appear at the periphery of each multiword sequence. For example, Table 5 shows *I want to* is the base structure of a common 4-word sequence in the lab corpus. Words that are likely to precede this structure are *so, what, because, is, and, now, that, all, and do*. Words that are likely to follow this structure are *change, have, select, find, reuse, use, get, count, put, and do*.

The majority of the clusters are made up of words from the 1st 1000 BNC/COCA frequency level. Some exceptions are: (I) know what the (energy/velocity), producing better value for, want a formula that, (I) want a function (that), in the public sector, the number of (distracters), and what the velocity (is). These clusters have items from the 2nd 1000 (energy, value), the 3rd 1000 (formula, function, sector), 4th 1000 (distract), and the 5th 1000 (velocity).

4. Discussion

The interview data concurred with the purposes of interactive speaking raised by Basturkmen (2016), including speaking to consolidate thinking and content knowledge, expressed here as a belief of the lecturing staff. Confidence and fluency in English were identified as core problems by the international students in this data set, along with difficulties following unfamiliarity with accents and rates of speech, cultural differences in approaches to speaking, and demands of frequent turntaking. It would be useful for students to be aware of lecturers' motivations behind

speaking in small groups at university as much as lecturers need to be aware of the difficulties of what Csomay (2006) calls real-time interaction.

The vocabulary load analysis found that tutorials and labs require fewer word families to reach 95% and 98% coverage of lectures and seminars (4000 word families and 8000 word families) (Dang and Webb (2014)), but are fairly similar to the vocabulary size necessary to reach 95% and 98% coverage of general spoken English (3000 word families and 6000 e7000 word families) (Nation, 2006; Webb & Rodgers, 2009a, 2009b). From a lexical coverage perspective, understanding labs and tutorials should not pose particular difficulties for L2 learners who have the level of English generally required for enrolment in an English-medium university. Academic talk is not full of technical lexis, as learners might expect (Dang et al. (2017)), and what learners read to prepare for small group classes on the same topic will not have the same lexical profile as the spoken text.

The analysis of the textbooks in this study suggests that these resources contain little in the way of focus on lexis in interactive speaking, particularly in laboratories. These findings are consistent with studies that compared the vocabulary in EAP textbooks and university textbooks (Chen, 2010, 2010; Miller, 2011; Wood & Appel, 2014) and echo the concern about the mismatch between EAP textbooks and information in corpora (Flowerdew & Miller, 1992; Harwood, 2005). A large proportion of the 176 patterns in three of the textbooks did not appear in the corpora. The multi-word units which did appear in the corpora (and not in the textbooks) are mostly made up of high frequency words, with some lower frequency words, and can be categorized into functions such as stance expressions like (what) I want to (do) and (I/we/you) don't know (how/what/I/ if), which might help L2 speakers with presenting ideas or turn taking.

5. Pedagogical and textbook implications

It seems clear that EAP learners need to be aware of the reasons why a lecturer at university might require them to speak in class, just as EAP teachers and university lecturers need to be aware of, and work to mitigate, the barriers for successful communication in tutorials and laboratories for L2 speakers of English. This research suggests that high frequency words, especially the first 1000 words, are vitally important for speaking in tutorials and labs. Teachers and learners could perhaps draw on research such as Dang et al. (2017) Academic Spoken Word List. Learners might think they already know these words, but in small group interactions, they may be used in fast turn taking, with colloquial meanings, as part of the subject knowledge or

part of a joke or aside, and their use might not be predictable. Teachers could provide learners with opportunities to encounter and use lexical bundles such as those identified in this study (see Appendix 4). Teachers could provide learners with examples from publicly available corpora (e.g., MICASE, HKCSE), examples from the tutorials or labs recorded by themselves, or through observing actual tutorials in university settings, and then help the learners to identify, analyse and practise the language patterns Basturkmen (2016). Learners also need opportunities to listen to speakers with different accents and interact in small group interactions to prepare for taking part in tutorials and laboratories at university. Finally, findings from this study could be considered by publishers and textbook writers.

6. Limitations and future research

This study has a number of limitations. The corpora for tutorials and labs are quite small and not very well balanced. Reviewers noted that Business makes up nearly half of the tutorials corpus, which raises the question of whether producing better value for is common in business discussions, common in small-group academic interactions, or common in both, and that variations in word order such as I agree completely and I completely agree have not been taken into account. The textbook selection could have been larger and the interviews did not directly focus on vocabulary in speaking. Suggestions for future research include a larger scale study of the vocabulary of tutorials and laboratories which could validate or confirm the vocabulary load and multi-word unit findings of this study. Cross-matching the multi-word units from this study with Simpson-Vlach and Ellis's (2010) academic formulas research would be useful. More research is needed on the vocabulary of different disciplines, for example, laboratories in Chemistry and Psychology. A longitudinal study of listening and speaking in tutorials and labs would shed light on vocabulary choices and usage (Sinclair & Renouf, 1988), development of fluency with high frequency vocabulary, and the effect of the local context on vocabulary use. See also Basturkmen (2016) for suggestions. Finally, tools other than corpus analysis may be better at shedding light on the pragmatic difficulties of speaking in small groups at university which are identified in this study. A reviewer suggested an approach such as conversation analysis of some of the tutorial and laboratory data (see McCarthy & Carter, 1997; Sinclair & Renouf, 1988) as a useful way forward.

7. Conclusion

This article has drawn on information from lectures, learners, corpora, and textbooks. It reveals

that engaging in small group speaking such as university labs and tutorials is not difficult for L2 learners from the perspective of lexical demands, yet learners reported it as a challenge. This study calls for more attention from learners, teachers, and course book writers on the lexical aspects of interactive dialogues in academic speaking, and provides corpus-based data on multi-word units which occur in tutorials and laboratories.

Table 1
Interview participants: Lecturers and students.

Participant	Applied Linguistics	Engineering	Business School	Total number
Lecturers	5	7	5	17
Students	8	8	8	24
Total	13	15	13	41

Table 2
Composition of the lab and tutorial corpora.

Labs: Subject areas	Number of words	Tutorials: Subject areas	Number of words
Biology	30,734	Art	11,303
Chemical engineering	18,739	Business	185,333
Chemistry	8061	Communication ^a	6075
Civil engineering	9962	Education	87,766
Communication ^a	8756	Engineering	38,638
Computer science	12,337	English	1014
Engineering	13,117	History	5606
Film, theater, music	5727	Political science	36,945
Mathematics	6209	Textiles & Clothing	7398
Psychology	23,757		
Total	137,399	Total	380,078

^a Communication had both tutorials and labs.

Table 3
Coverage of Nation's (2012) BNC/COCA lists in the laboratory and tutorial corpora.

Word list	Coverage at each level		Cumulative coverage (plus proper nouns & marginal words)	
	Laboratories	Tutorials	Laboratories	Tutorials
1 st 1000	85.46	86.54	89.55	89.96
2 nd 1000	4.26	4.62	93.81	94.58
3 rd 1000	2.22	3.19	96 ^a	97.77 ^a
4 th 1000	0.89	0.68	96.92	98.45 ^b
5 th 1000	0.44	0.28	97.36	98.73
6 th 1000	0.44	0.19	97.8	98.92
7 th 1000	0.2	0.13	98.0 ^b	99.05
8 th 1000	0.17	0.08	98.17	99.13
9 th 1000	0.16	0.06	98.33	99.19
10 th 1000	0.13	0.04	98.46	99.23
11 th -25th 1000	0.67	0.16	99.13	99.39
Proper Nouns	1.57	1.06		
Marginal words	2.52	2.36		
Compound words	0.28	0.18		
Not in any lists	0.37	0.25		

^a Reaching 95% coverage.

^b Reaching 98% coverage.

Table 4

Frequency per million of the 176 patterns in the lab and tutorial corpora.

Relative frequency	Labs	Tutorials	Labs & tutorials	Examples from corpora
0	130	114	111	you're kidding, I disagree with
1-9	5	22	27	agreed that
10-19	5	11	7	I agree with, that's interesting
20-29	4	2	2	that's true, excuse-me
30-39	4	2	3	what do you mean
40-49	2	3	5	I agree
50-59	5	5	2	do you mean, how about
60-69	2	0	0	I believe
70-79	0	0	1	thought that
80-89	4	3	4	what do you think, I see
90-99	0	2	0	what about
100-999	9	6	8	In other words, Wow, thank you
1000 and above	6	6	6	okay, yeah, I think

Table 5

Examples of the clusters appearing in the lab corpus.

Multi-word units	Freq/million
Stance expression	
(so/what/because/is/and/now/that/all/do) I want to (change/have/select/find/reuse/use/get/count/put/do)	1033
(we/you) don't know (how/I/if/is/it/what/why)	655
Discourse organizers	
I think that (s)	109
(if) you look at (the)	102
Referential expression	
(but) say for example (you/I)	116
the rest of (the/of)	109
Special conversational functions	
What's going on	51

Table 6

Examples of the clusters appearing in the tutorial corpus.

Cluster	Freq/million
Stance expression	
(I/we/you) don't know (how/what/I/if)	526
(what) I want to (do)	66
Discourse organizers	
(if) you look at (the)	158
I think that (s)	97
Referential expression	
That's what (I/we/it)	132
(all) the rest of (it/the)	103
Special conversational functions	
What's going on	92

Table 7

Examples of the clusters appearing in the corpus of labs and tutorials.

Cluster	Freq/million
Stance expression	
(I/we/you) don't know (what/if/I/how)	773
(what/if/so) I want to (do)	186
Discourse organizers	
(if) you look at (the/it)	172
I think that (s)	100
Referential expression	
(and) that's what (I/it/we/you)	240
the rest of (the/it)	73
Special conversational functions	
What's going on	81

References

Author (2000)

Authors (2002)

Authors (2007)

Authors (2008)

Authors (2014)

Authors (2017)

Ackermann, K. & Chen, Y-H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247.

Basturkmen, H. (2016). Dialogic interaction. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 152–164). London: Routledge.

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6: 253–279.

Becher, T. (1989). *Academic tribes and territories*. Bristol, UK: The Society for Research into Higher Education and Open University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing.

- Biber, D., Conrad, S. & Cortes, V. (2004). If you look at...: Lexical bundles in University teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
- Bondi, M. (2016). Textbooks. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 323-334). London: Routledge.
- Chen, L. (2010). An investigation of lexical bundles in ESP textbooks and electrical engineering introductory textbooks. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 107–125). London: Continuum.
- Corbin, J. M. & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). London: SAGE Publications.
- Crawford Camiciottoli, B. (2010). Meeting the challenges of European student mobility: Preparing Italian Erasmus students for business lectures in English. *English for Specific Purposes*, 29, 268–280.
- Csomay, E. (2006). Academic talk in American university classrooms: Crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes*, 5(2), 117–135.
- Feak, C. (2013). ESP and speaking. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 34-53). Oxford: Wiley-Blackwell.
- Flowerdew, J. & Miller, L. (1992). Student perceptions, problems and strategies in second language lecture comprehension. *RELC Journal*, 23(2), 60–80.
- Gardner, D. and M. Davies. (2014). A new academic vocabulary list, *Applied Linguistics* 35(3), 305–327.
- Glaser, B. G. & Strauss, A. L. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes*, 4, 149–161.
- Hincks, R. (2003). Pronouncing the Academic Word List: Features of L2 student oral presentations. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetics Sciences* (pp. 1545–1548). Barcelona, Spain. Retrieved from http://www.speech.kth.se/ctt/publications/papers03/icphs03_1545.pdf
- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge, UK: Cambridge University Press.

- Kelly-Laubscher, R., Muna, N. & van der Merwe, M. (2017). Using the research article as a model for teaching laboratory report writing provides opportunities for development of genre awareness and adoption of new literacy practices. *English for Specific Purposes*, 48, 1–16.
- Kim, S. (2006). Academic oral communication needs of East Asian international graduate students in non-science and non-engineering fields. *English for Specific Purposes*, 25, 479–489.
- Lei, L. & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42–53.
- Marlina, R. (2009). “I don’t talk or I decide not to talk? Is it my culture?”—International students’ experiences of tutorial participation. *International Journal of Educational Research*, 48, 235–244.
- Miller, D. (2011). ESL reading textbooks vs. university textbooks: Are we giving our students the input they may need? *Journal of English for Academic Purposes*, 10(1), 32–46.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82.
- Nation, I. S. P. (2012). The BNC/COCA word family lists. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P. & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle, Cengage Learning.
- Nesi, H. (2002). An English Spoken Academic Word List. In A. Braasch & C. Povlsen (Eds.), *Proceedings of the Tenth EURALEX International Congress (Vol. 1, pp. 351–358)*. Copenhagen, Denmark. Retrieved from http://www.euralex.org/elx_proceedings/Euralex2002/036_2002_V1_Hilary%20Nesi_An%20English%20Spoken%20Academic%20Wordlist.pdf
- Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, 26(2), 135–146.
- Paribakht, T. S. & Webb, S. (2016). The relationship between academic vocabulary coverage and scores on a standardized English proficiency test. *English for Academic Purposes*, 21, 121–132.
- Parkinson, J. (2017). The student laboratory report genre: A genre analysis. *English for Specific Purposes*, 45:1–13.

- Reinhart, S. (2002). *Giving academic presentations*. Ann Arbor, MI: University of Michigan Press.
- Schmitt, N., Cobb, T., Horst, M. & Schmitt, D. (2015). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*. <https://doi.org/10.1017/S0261444815000075>
- Simpson-Vlach, R. & Ellis, N. C. (2010). An Academic Formulas List: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512.
- Skyrme, G. (2010). Is this a stupid question? International undergraduate students seeking help from teachers during office hours. *Journal of English for Academic Purposes*, 9, 211–221.
- Tapper, J. (1994). Directives used in college laboratory discourse. *English for Specific Purposes*. 13(3), 205–222.
- Thompson, P. (2006). A corpus perspective on the lexis of lectures, with a focus on economics lectures. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 253–270). New York: Peter Lang.
- van Zeeland, H. & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479.
- Watson-Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes*, 45, 31–39
- Webb, S. & Paribakht, T. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes*, 38, 34–43.
- Webb, S. & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407–427.
- Webb, S. & Rodgers, M. P. H. (2009b). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366.
- Wood, D. (2010). Lexical clusters in an EAP textbook corpus. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 88–106). London: Continuum.
- Wood, D. C. & Appel, R. (2014). Multiword constructions in first year business and engineering university textbooks and EAP textbooks. *Journal of English for Academic Purposes*, 15, 1–13.
- Yang, M-N. (2015). A nursing academic word list. *English for Specific Purposes*, 37, 27–38.

Appendix One: Interview guides for lecturers and students

Lecturer interview guide

1. Activity identification: What sorts of classroom activities do you choose for this particular course?
2. Academic level: Do you set up different activities between 400/500 (postgraduate) and 300 and below (undergraduate)?
3. Course nature: Do you think you set up different activities for your different courses?
4. Active learning: How do you expect students to participate in classroom activities?
5. Collaborative learning activities: Do you give collaborative learning activities in or outside the classroom? If you do, why and what do you expect from that? Any issues?
6. Assessment: Do you assess students' participation in classroom activities?
7. Personal belief behind practices: What influenced your current style of teaching practice? Disciplinary expectation/constraint, personal experience as students and teachers, career development resources such as books, or personal experiences in the real world?
8. NS-NNS interaction: Do you find any issues with your interaction with NNS students or peer interaction between NS and NNS students?

Student interview guide

1. Activity type and study level: Do you find any difference in classroom activities between your current postgraduate studies and previous undergraduate studies?
2. Intercultural difference: Do you identify any 'Kiwi' culture in terms of classroom learning activities and classroom interaction?
3. Language: Do you find any difference between your own first language (L1) and second language (L2) use when you participate in classroom interaction? Does your L2 cause any specific difficulty in the academic interaction?
4. Interaction type: Do you voluntarily give questions or comments to lecturers in lecture? Do you think you behave differently or feel different when you talk with lecturers and peer classmates? How do you participate in the talk held between the lecturer and another student? Do you think you take the floor voluntarily during a talk held between the lecturer and another student? If you do, when do you think you need to do so?

5. Peer talk/work: What do you think is the point of doing peer talk/work in classroom? Do you think it benefits you in any way? Do you find any issue with peer interaction? If you do, how do you cope with that?
6. Native Speakers VS Non-Native Speakers: Do you find any difference between NS and NNS in terms of participation and behaviour in classroom interaction? Do you find any benefit or issue with interaction with NS/NNS?

Appendix Two: Names and publishers of EAP and ESP textbooks used in this study

	Series	Publishers
EAP textbooks	Academic listening strategies: a guide to understand lectures	University of Michigan Press
	Communicating on campus	Altar book center
	Contemporary topics	Pearson Longman
	EAP Now	Pearson Longman
	English for academic study (Listening)	Garnet Education
	Intermediate listening comprehension	Heinle-Cengage ELT
	Learning to listen, listen to learn 1	Pearson Longman
	Lectures: Learn listening and note-taking skills	Collins EAP
	Lecture ready	Oxford University Press
	Listening & Notetaking Skills 3 Student Book	
	Advanced Listen	Heinle-Cengage ELT
	Listening power	Pearson Longman
	Oxford EAP: A course in English for academic Purposes	Oxford University Press
	Passport to academic presentation	Garnet Education
	Study Listening: A Course in Listening to Lectures and Note Taking	Cambridge University Press
	ESP textbooks	English for Business Studies in Higher Education Studies
English for Economics in Higher Education Studies		Garnet Education
English for Environmental Science in Higher Education Studies		Garnet Education
English for ICT Studies in Higher Education Studies		Garnet Education
English for Languages and Linguistics in Higher Education Studies		Garnet Education
English for Law in Higher Education Studies		Garnet Education
English for Management Studies in Higher Education Studies		Garnet Education

English for Mechanical Engineering in Higher Education Studies	Garnet Education
English for Psychology in Higher Education Studies	Garnet Education
English for Tourism and Hospitality in Higher Education Studies	Garnet Education
English for Public Relations in Higher Education Studies	Garnet Education

Appendix Three: Categorisations of the 176 phrases from Oxford EAP, Communicating on Campus, EAP Now

1. Expressing agreeing/ Disagreeing (e.g., I simply disagree, I totally agree)
2. Checking for comprehension/ asking for clarification/ confirmation/ elaboration (e.g., Could you please say that a little more slowly/loudly, Could/can you clarify)
3. Giving opinions (e.g., I think, in my experience/ opinion, as far as I'm concerned).
4. Responding (express interest/ surprise) (e.g., hmm, that's interesting, you're kidding)
5. Offering a fact or example (e.g., let me give you an example, I'd like to add)
6. Asking for opinions/ ideas (e.g., what's your take on that, what's your opinion of)
7. Keeping the discussion on topics (e.g., let's get back to, we are getting a little off track).
8. Expressing compromise and Reaching consensus (e.g., in our discussion we talked about, we came to the conclusion that)
9. Giving advice (e.g., if I were you I'd, you may want to)
10. Asking for advice (e.g., what do you think/suggest I should do, what should I do)
11. Giving explanations (the reason is)
12. Introducing the topic (e.g., let's start with, we need to discuss)
13. Paraphrasing (e.g., in other words, what she meant was)
14. Taking time to brainstorming the ideas (let me think about that for a moment)
15. Expressing uncertainty (that's a very interesting question but I'm afraid I don't know, I'm not sure, but I'll try to find out for you)