THE JOURNAL OF ASIA TEFL
Vol. 17, No. 1, Spring 2020, 70-86
http://dx.doi.org/10.18823/asiatefl.2020.17.1.5.70

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

Lexical Richness of One-Minute Speaking Task by Science and Technology University Students

Thiwaporn Thawarom
King Mongkut's University of Technology Thonburi, Thailand

Wareesiri Singhasiri
King Mongkut's University of Technology Thonburi, Thailand

Driven by Coxhead's question "What vocabulary do ESP learners need?", this study aims to investigate the quality of vocabulary use by measuring lexical richness (lexical diversity, lexical sophistication and lexical density) of speech production in ESP classroom context. The data was collected from 100 first year university students while performing a speaking task. Drawing on D_tools and AntWordProfiler, we made measures of diversity and sophistication. TagAnt was applied for calculating percentages of lexical density. Furthermore, a scale-based approach was employed for technical identification. The results revealed that vocabulary produced by the students was mainly from GSL1000 word list. Lexical density of the speech was 43.72% and the lexical diversity was moderately rich with a value of D 58.8. This study suggests needs of productive vocabulary in ESP classroom contexts. The results of this study imply that ESP students need a great awareness that they are parts of a particular group. To achieve the goal of learning, they are required possession of disciplinary knowledge and skills of the area they belong.

Keywords: lexical richness, lexical diversity, lexical sophistication, lexical density, ESP speaking

Introduction

Vocabulary is a key component for language use and a good indicator of mastering second language performance (Schmitt, 2010). Learners of a second language need to acquire knowledge of vocabulary for learning as well as for communication (Akbarian, 2010; Goh, 2007; Li & Wai Chum, 2012). For example, learners should know at least 95 per cent of the words in a reading text in order to comprehend (Laufer, 1989) and 90 per cent of words for comprehension of a spoken text (Van Zeeland & Schmitt, 2013). However, a question arises when the context is discipline-specific, where specialized vocabulary belongs to a particular subject area. According to Coxhead (2013, p. 116), the size of specialized vocabulary has not fully established. Despite the fact that, the size depends on subject areas (Nation, 2008). In this sphere, learners need to build up knowledge of the technical vocabulary that they are likely to encounter in order to master disciplinary knowledge.

Studies of ESP vocabulary have placed priority on vocabulary needed for receptive skills. Previous studies included technical vocabulary in anatomy textbooks (Chung & Nation, 2004), science textbook (Veenstra & Sato, 2018), engineering textbooks (Hsu, 2014; Jin et al., 2013; Ward, 2009; Watson Todd, 2017), business studies lectures (Crawford Camiciottoli, 2007), financial studies (Kwary, 2011), and

plumbing workbooks, unit standards and instruction manuals (Coxhead & Demecheleer, 2018). Lists of technical vocabulary are believed to enrich the students' reading and listening comprehensions. However, the effectiveness of such scaffolding is unclear unless the students' comprehension is tested. To do so, an assessment such as a speaking task allows teachers to gauge whether ESP students are capable of using vocabulary and language successfully and purposefully. Thus, investigating productive vocabulary in an ESP context may shed light on the vocabulary study.

Specifically focusing on an ESP classroom context, this paper investigates vocabulary richness of one-minute speaking performed by EFL science and technology students. The context of the present study is at a science and technology university, King Mongkut's University of Technology Thonburi (KMUTT) - Thailand. The university excels in architecture, Bioresources, energy, engineering, environment, linguistics, and science and technology. Apart from content subjects, the English curriculum for the undergraduate students emphasizes on academic English and technical English via skill-based and task-based approaches. Among all the English courses, Technical English is a course, which aims to provide students with English skills for everyday communication at works including technical, industrial, and scientific sectors. Hence, the students are expected to gain general technical vocabulary, retain and pertain them purposefully and appropriately.

A common ground from the previous ESP studies lies in the fact that vocabulary is generally conceived of as a scaffold for the learners. In fact, the quality of vocabulary can have an impact on task achievement, so there is a necessity to examine vocabulary use or lexical richness as the output of the learners can predict their speaking task performance. Therefore, this study seeks to answer the following question.

> What is the lexical richness – diversity, sophistication, and density – of EFL science and technology students in performing a speaking task?

Literature Review

English for Specific Purposes (ESP)

English for Specific Purposes (ESP) refers to the teaching and learning of English as a second or foreign language where the goal of the learners is to use English in a particular domain (Paltridge & Satarfield, 2013, p. 2). Originally, ESP emerged from the demand of international language for communication in technology and commerce. Afterwards, demands for English to suit particular needs and developments in various fields arose (Huchinson & Waters, 1987). Subsequently, the approach has been expanded to other areas including English for Academic Purposes (EAP) (Benesch, 2001). According to Flowerdew and Peacock (2001), ESP is an umbrella term that covers two main categories i.e. English for Academic Purposes and English for Occupational Purposes. English for Academic Purposes ranges from English for Business to English for Science and Technology. Whereas, English for Occupational Purposes covers English for professional purposes such as vocationally oriented courses.

ESP courses focus on language, skills and genres appropriated to the specific activities that the learners need to carry out in English (Paltridge & Satarfield, 2013). Thus, ESP materials, activities and assessments are imperative to be implemented in classroom practices. As an example, a task-based activity in an ESP classroom might assign students to do a technological project, which allows them to extend their language use to a real-word situation. Accordingly, learners at this level are anticipated to be competent at integrating language and content knowledge (Douglas, 2000). Furthermore, requisite knowledge of particular subjects and applications are significant for future careers (Coxhead, 2013; Douglas, 2001; Fiorito, 2005).

ESP Vocabulary and Speaking

ESP vocabulary refers to the vocabulary of a particular area or professional use. The term is denoted with different names in the literature including special purpose vocabulary, specialized vocabulary, technical vocabulary, sub-technical vocabulary, and semi-technical vocabulary (Coxhead, 2013, pp. 115-116). Technical vocabulary can be characterized by having a very specific and not possessing direct synonym (Mudraya, 2006). This makes ESP vocabulary very difficult for students to comprehend or guess because of its narrow range (Nation, 2001). ESP vocabulary, however, does not only include long Greco- Latin words or highly technical words that are not used in everyday language, but also everyday words that embrace very specific meanings in particular contexts (Coxhead, 2013). For instance, *market* and *price* are both commonly used in everyday situation, but they are also used with precise technical meanings in business studies (Crawford Camiciottoli, 2007). Hence, it is difficult to tell where technical vocabulary starts. However, ESP vocabulary can be identified by different approaches such as consulting experts (Schmitt, 2010), using a scale (Chung & Nation, 2003) and using a common core approach (Coxhead, 2000).

In speaking, approximately 95% of words are obtained from 2,000-word families (Adolphs & Schmitt, 2003). When speaking in ESP context, there are necessities for learners to combine both general words and technical words or words beyond 2000. Accordingly, speakers may find speaking difficult because of limited vocabulary from their lexicon (Nation, 2013; Zipagan & Lee, 2018). Furthermore, situations and variety of topics shape vocabulary they will produce (Richards, 1976). Possibly, speaking is demanding due to technical-oriented words to be used. Thus, exploring vocabulary use or lexical richness in technical contexts may disclose another window of ESP vocabulary studies.

Lexical Richness

Lexical richness can be an indicator of the language learners' proficiency of the written and spoken productions. In speaking, the quantitative lexical richness of a speaker refers to the amount of vocabulary that the speaker freely uses in discourse (Zhang, 2014, p. 61). According to Read (2000, p. 200), lexical richness is a general term used for measuring characteristics of effective vocabulary use. It is multifaceted, consisting of four dimensions comprising lexical diversity, lexical sophistication, lexical density and numbers of errors. However, confusion may arise as different researchers define and conceptualize the term lexical richness differently. Some researchers consider lexical richness to be interchangeable with lexical diversity (Daller, Van Hout, & Treffers-Daller, 2003; Foster, 2009; Malvern & Richards, 2002; Torruella & Capsada, 2013). The current study sides with Read (2000), which acknowledges the values of different aspects of lexical richness. The study focuses on three of the four aspects – diversity, lexical sophistication, and lexical density. The following sections provide details of each of these three dimensions.

Lexical diversity

Lexical diversity can indicate the language learners' proficiency of maximizing different words in written or spoken texts. Learners of those of higher proficiency are more likely to use diverse vocabulary and avoid vocabulary repetition. Johansson (2008) found that lexical diversity could detect the differences between age groups in a spoken text. That is, adults and 17-year old groups produced a more varied vocabulary in narrative spoken texts than younger age groups (13-year olds and 10-year olds). Lu (2012) studied the relationship of lexical richness in three aspects including variety, sophistication and density against the quality of ESL learners' oral narratives. It was found that among these aspects, lexical diversity had the strongest correlation with the rater's judgment of the quality of the oral narrative of ESL learners.

Traditionally, lexical diversity is measured as a Type-Token Ratio (TTR), which is the number of different words (types) divided by the total number of all words (tokens) produced in a spoken or a written text. This variable shows that the more varied vocabulary presents in a text, the greater the lexical diversity is. However, when the text length increases, words tend to be used repeatedly, and new words will be less likely to occur (McCarthy & Jarvis, 2007). Moreover, when texts are different in lengths, TTRs do not yield consistent results. Due to the limitation of its sensitivity to the text lengths, other measures of TTR with different variants have been proposed. These include RTTR (root type-token ratio) (Guiraud, 1954) and CTTR (corrected type-token ratio) (Carrol, 1964), Herdan (H) index (Herdan, 1960), D (Malvern & Richards, 1997), MTLD (measure of textual lexical diversity) (McCarthy, 2005), HD-D (McCarthy & Jarvis, 2007) and MTLD (McCarthy & Jarvis, 2010).

Lexical sophistication

Lexical sophistication refers to low-frequency vocabulary that are appropriate to a specific topic and style of writing, rather than simply general, everyday vocabulary. This includes the use of technical terms and jargons to express meanings in a precise and sophisticated manner (Read, 2000, p. 200). The amount of low-frequency words that language learners use in their writing or speaking affects their holistic scores and/or proficiency scores when judged by human raters (Kyle & Crossley, 2015).

Lexical sophistication in second language is commonly based on frequency bands (Luafer & Nation, 1995). To be exact, word families (danger, dangerous, dangerously) are categorized into frequency bands of the most common 1,000, 2,000, 3000, and so on. Lexical Frequency Profile, LFP (Luafer & Nation, 1995), which measures a proportion of high frequency and low-frequency words in a text can be applied to indicate a level of lexical sophistication. Vocabprofile is a program that computes the General Service List of English Words (GSL) (West, 1953) where words that are of high frequency in most uses of the language, to Academic Word List (AWL) (Coxhead, 2000). This means low-frequency words that rarely occur in general contexts are considered more sophisticated. These words tend to appear a narrow range of use within particular subject areas such as the use of words in academic fields and in technical contexts.

For technical oriented words, different approaches of identification have been employed. Jin et al. (2013) drew on engineering technology textbooks as a corpus to compare with the larger reference corpus BNC to help discern technical words. Watson Todd (2017) applied the ‘opaque’ concept, which dealt with words that have unusual meanings for the identification. Additionally, Kwary (2011) recognized technical vocabulary in a Chartered Financial Analyst textbook by utilizing a hybrid method by evaluating merits and demerits of four methods. These included vocabulary classification, keywords analysis, term extraction and systematic classifications. Combining keyword analysis method and the systematic classification method together, technical vocabulary was meticulously identified.

Furthermore, Chung and Nation (2003) employed a scale-approach to categorize technical vocabulary in anatomy and linguistic texts. The scale focuses on strength of the relationship of a word to a specialized field, which in their example case was anatomy. The scale consists of four steps. Step 1 represents words with meanings that have no particular relationship with the field of anatomy. Step 2 embodies words having meanings that are minimally related to the field of anatomy. Step 3 contains words that embrace a meaning that is closely related to the field of anatomy. Step 4 is for words that have a meaning specific to the field of anatomy. Words from Step 3 and 4 were considered technical. They reported that the learners were likely to encounter one technical vocabulary in every three running word in the anatomy text.

Lexical density

Lexical density measures the lexical or content words in a text. It has been employed for indicating the quality of learners’ writing (Laufer & Nation, 1995) and speaking performances (Jarvis 2002; Malvern &

Richards, 2002). Lexical density is the only feature of lexical richness that distinguishes written from spoken language (Didau, 2013; Read, 2000; Ure, 1971). According to Ure (1971), written texts are usually comprised of more than 40 percent lexical or content words, whereas spoken texts generally consist of less than 40 percent. Regardless of these percentages, lexical density plays a vital role in both writing and speaking. Precisely, the writing becomes too simple (To, Fan, & Thomas, 2013), or meaningless and vague if the number of lexical words is too low (Didau, 2013). In an academic register, a speaker needs to convey information in an 'explicit manner' in order to provide the listeners with enough information about the specific topic (Henrichs & Schoonen, 2009).

To measure lexical density, content words (nouns, verbs, adjectives, and some adverbs) are compared to the total number of words (Ure, 1971). Nonetheless, 'content word' can be conceptualized differently. Halliday (1985), counts lexical items, which may consist of more than one word. For example, *turn up* is counted as one lexical items. On the contrary, Ure considered *turn* as lexical item and *up* as grammatical item. This current paper, we follow Ure's (1971) definition.

From the literature, there are two reasons to extend the study of lexical richness. Firstly, studies of lexical richness of learners' speech production of ESP in classroom contexts have not been uncovered. Previous studies have focused on lexical richness of speech production in general contexts (narrative and expository) (Johansson, 2008), academic contexts (Dang & Webb, 2014; Kyle & Crossley, 2015). While, some studies such as Nation and Chung (2004), Ward (2009), Coxhead and Demecheleer (2018) spotted on lexical richness in written texts in technical contexts. Another reason is that investigation of lexical richness of EFL learners may contribute more insights of vocabulary studies. In previous studies, lexical richness as the synonym of lexical diversity treated rarity of words in measurement. In this current study, lexical richness is perceived to be multi-facet and treats each facet differently. We attempt to find an approach to show the learners' capability of generating or exploiting appropriate vocabulary to the context given. In order to evaluate the learners' speaking, one facet might be more essential than the other might. To do so, we provide the methodology in the following section.

Methods

Participants

The participants in this study were 100 first-year science and technology undergraduate students who enrolled in the course Technical English in Semester 1 of Academic Year 2017. They were exempt from the first fundamental English course due to the scores of their placement test, which was Ordinary National Education Test. Thus, their English proficiency was assumed to be equivalent to an intermediate level. The participants were selected purposively according to their availability for data collection. In total, there were three cohort classes (31, 32 and 37 each). The students in each class were mixed in terms of their majors. They were from eight departments: Mechanical Engineering, Mechatronics Engineering, Electrical Engineering, Control System and Instrumentation Engineering, Microbiology, Media Arts, Media Technology, and Medical and Science Media.

Speaking Task

Technical English is a course that aims to teach essential technical communication skills for engineers, technicians and scientists to students. The students are expected to be able to communicate in particular ESP contexts such as in laboratories, factories, and construction sites. The main language skills taught in this course are speaking and writing. Therefore, the requirements of tasks focus on productive skills such as describing situations, writing instructions, writing technical report and giving presentation.

One of the speaking tasks, which was selected for this study, was a one-minute speaking. The theme of the task was safety in the workplace. The students were required to show their ability to orally describe a

situation about an accident in the workplace. They were provided with guidelines to describe the scene, equipment or materials and potential dangers, which appeared in a picture for one minute. The task's guidelines were (1) What happened in the scene?, (2) How serious was the case?, (3) What might be the cause?, (4) What are the equipment involved?, and (5) What are correct positions of things or equipment in the scene? There were five different pictures. When performing the task, one by one, the student randomly picked one picture before giving a description of that picture. While they were performing the task, the students' speech was audio recorded.

The pictures of accidental situations for the speaking task were available on the websites, which allow users to have access for educational purposes (See index).

Technical Supports

In order to examine lexical richness of EFL learners, three automated programs were employed as follows:

1. D_Tools (Meara & Miralpeix, 2015) is an online tool for measuring diversity based on work by Malvern et al. (2004). The program works by taking a set of 100 samples of 35 words each from the text and computes a mean Type-Token value for each of these samples. The program will then take 100 samples of 36 words and computes the mean TTR for these samples. Samples of 37 words, 38 words, 39 words, and so on up to 50 words are taken, and the mean TTR of each set of samples is computed. The means of TTR will then be compared to the best-fitting value of D in Malvern and Richards' theoretical model. The tool also reports an error score, which indicates the closeness of model data and the actual data. If the error score is more than 0.01, the model is not a good match for the data. This tool is available at http://www. lognostics.co.uk/tools/ D_Tools/D_Tools.html.
2. AntWordProfiler (Anthony, 2014), a modern version of RANGE, is a program to analyze vocabulary load in a text. It allows users to generate a profile of vocabulary in a text by comparing the vocabulary in the text against default vocabulary lists, which include the GSL1000, GSL2000 (West, 1953) and AWL (Coxhead, 2000). The output displays statistics of vocabulary and frequency information about the text according to the level (default) lists, plus words that do not belong to the default lists. Therefore, this program measures the lexical sophistication of the text. AntWordProfiler is freeware software available at http://www.laurenceanthony.net/ software.html.
3. TagAnt (Anthony, 2015), is a tagging tool based on TreeTagger. The tool assigns a Part-Of-Speech (POS) Tagger to each word in a text file. POS taggers allow users to see both content and functional words in a text. In the current study, it was applied to calculate percentages of lexical density produced by the EFL students. TagAnt is also freeware software and is available at http://www.laurenceanthony.net/software.html.

Procedures

Preparation for transcripts

1) When the speaking task ended, the speaking audios were transcribed utilizing verbatim transcription. That is, everything the researchers heard in the audio recording was transcribed (Paulus, Lester, & Dempster, 2014 p. 96). However, words that were unintelligible by the agreements of the researchers and the two experts due to poor articulation were dismissed from the transcripts.

2) The transcripts were cleaned according to the following criteria:
   2.1) As speaking lasted only 1 minute, the part that exceeded 1 minute was omitted. After that, miss-spellings/ miss-typed were rechecked.
   2.2) Words produced in the students' first language (Thai) while speaking were excluded as the focus of this study was to examine the use of vocabulary in English. Thai words and phrases, therefore; were discarded.
   2.3) Unfinished and mispronounced words were rejected due to incomplete expressions and meanings.
   2.4) Words that lacked senses of grammatical or contextual meanings (hesitations) i.e., *um*, *ah*. Sounds that the speaker made when he or she was thinking about what to say next were also removed.
3) After the cleaning process, the transcripts were converted into text files for the compatibility of the software programs for analysis.

Running for lexical diversity

Process the cleaned a text file(s) into the D_Tools application by pasting them into a *Textbox*. Then, name the file and click *Submit*. The program shows the value of D in a figure with two decimal points. The means' TTRs of the set of samples compared with the value of the model data are shown in a graph. The error score is reported too.

Running for lexical Sophistication

The text file(s) were processed into the automated program, AntWordProfiler, for profiling vocabulary levels by clicking *File >Choose User File (s)*. Under the *Output Settings*, click the boxes to choose the outputs and click *Start*. The program gives the numbers of tokens, types, and groups (families) as well as a percentage of each vocabulary level i.e., GSL 1000/ GSL 2000/ AWL and Non-listed. Moreover, the program provides the list of vocabulary of each level. Therefore, the lexical sophistication can be explored according to the list provided.

Running for lexical Density

Firstly, process the text file (s) into the TagAnt program for part of speech tagging by selecting *File>Open File (*s), then, select the *Language*. Last, choose the orientation of the output results by selecting either *Vertical* or *Horizontal*. Then, the researcher counted content words and calculated the percentage of lexical density. For the lexical density of each level, the list of vocabulary of each level was used as a benchmark to count POS tags. After that, the numbers of content words in each vocabulary level were counted and the lexical density of each level was calculated.

Identifying Technical vocabulary

Since the AntWordProfiler displays Non-listed words, which are neither general nor academic vocabulary, words appeared at this level are assumed to be technical. Thus, we applied the scale from Nation and Chung (2003) to help identify the technical vocabulary. Because of the vocabulary use in the context of these study is somewhat specific, the original scale is necessarily adapted. In the table below, it can be seen that only Steps 1 to 3 were modified because of the strength of the relationship of a word to the theme of the task.

TABLE 1
Scales for Identifying Technical Words Adapted from Chung and Nation (2003, 105)

Chung and Nation's scales	Modified scales
Step 1: Words such function words that have a meaning that has no particular relationship with the field of anatomy. Examples are the, is between, it, by, I2, adjacent, amounts, common, commonly, directly, constantly, early, and especially.	Step 1 Words that have a meaning that is minimally related to situation in that they describe the position, actions, places and objects
Step 2: Words that have a meaning that is minimally related to the field of anatomy in that they describe the position, movements, or features of the body. Examples are: superior, part, forms, pairs, structures, surrounds, supports, associated, lodges, and protects	Step 2 Words that have a meaning that is related to the situation which can explain the causes or effects of the situations
Step 3: Words that have a meaning that is closely related to the field of anatomy but are also used in general language, or may occur with the same meaning in other fields and not be technical terms in those fields. Examples are: chest, trunk, neck, abdomen, ribs, breast, cage, cavity, shoulder, girdle, skin, muscles, wall, heart, lungs, organs, liver, bony, abdominal, and breathing	Step 3 Words that have a meaning that is closely related to the situation and could be counted as technical terms in particular fields.
Step 4: Words that have a meaning specific to the fields of anatomy and are not likely to be known in general language. They refer to structures and functions of the body. These words have clear restrictions of usage depending on the subject fields. Examples are: thorax, sternum, costal, vertebrae, pectoral, fascia, trachea, mammary, periosteum, hematopoietic, pectoralis, viscera, intervertebral, demifacets, and pedicle	-

Results

The results consist of three parts: lexical sophistication, lexical density and lexical diversity.

Lexical Sophistication

This section reports lexical richness from the perspective of lexical sophistication. There are two parts: vocabulary profiles or levels of vocabulary use and technical vocabulary.

Vocabulary profiles

Vocabulary profile consists of three default level lists: GSL_1st1000, GSL_2nd 1000, and AWL_570 and one non-based list that represents words not belong to the default lists. Non-based is an optional output. As displayed in Table 2, the results highlighted that the students employed 7,884 tokens from GSL1000 when describing the situation. 596 tokens of GLS 2000 were produced, whereas, only 236 tokens from the AWL appeared. Of the total tokens, there were 504 types of vocabulary performed at GSL1000, 148 types at GSL 2000 and 43 types of AWL. This indicates that the students repeated vocabulary in their oral report. For word families, 351 belonged to GSL 1,000. Surprisingly, the students were able to use 118 word families at GSL2000. At the same time, the capability of producing AWL words was only 34 families.

TABLE 2
Lexical Sophistication in One-minute Speaking by Science and Technology Students

LEVEL	FILE	TOKEN	TOKEN%	TYPE	TYPE %	GROUP	GROUP%
1	1_gsl_1st_1000.txt	7884	85.17	504	60.87	351	55.19
2	2_gsl_2nd_1000.txt	596	6.44	148	17.87	118	18.55
3	3_awl_570.txt	231	2.5	43	5.19	34	5.35
0	-	546	5.9	133	16.06	133	20.91
TOTAL		9257		828		636	

Nevertheless, due to the nature of the task, the presence of technical words was expected. The vocabulary from the default lists have not yielded technical vocabulary use by the students. Interestingly, we found 546 tokens appeared in non-based list (Table 2). The number was almost double the tokens of AWL. These words were assumed technical because they were neither from GSL nor AWL lists. Furthermore, we modified Chung and Nation' (2003) scale to identified technicality of the words in non-based list.

Technical words

In order to identify technical vocabulary employed in the students' speech production, the closeness of relationship of a word to the theme of the task was applied. As seen in Table 3, the words which belonged to Step 1 such as *aisle*, *boss*, *carelessness*, *cartoon*, *clip*, *counter*, *guy*, *donut (s)*, and *chill* were used to describe the situation using general information. They were employed to describe the presence of objects in the picture (donut, microwave, teapot, and syrup). Words in Step 2 clarified the situation. Words such as *alarm*, *drain*, *pollute*, *stumble*, and *unorganized* portrayed more details about the situation and objects in the pictures. Whereas, words in Step 3 stipulated meanings that embraced the context. Some examples from this step were *scaffold*, *laboratory*, *radioactiv*e, *hazard*, *evaporate*, *contaminate*, and *fatal*. Thus, the words in this Step were considered to be technical. In this modification, no words belonged to Step 4 because the given situation was not highly technical that needs very narrow meaning words for description.

TABLE 3
Technical Vocabulary in a Speaking Task

Step 1 Words that have a meaning that is minimally related to situation in that they describe the position, actions, places and objects	Noun: aisle, boss, carelessness, cartoon, clip, counter, guy, donut (s), hub, jigsaw, microwave, movie, pancake (s), photo, pic, retriever, syrup, teapot, walkway Verb: chill, grasp, shorted, flipping, gonna, panicking, lain, retrieve Adj.: kinder (kind), sarcastic, thoughtless Adv.: anytime
Step 2 Words that have a meaning that is related to the situation which can explain the causes or effects of the situations	Noun: alarm, ban, chamber, craftsman, drain, drainpipe, duct, emergency, glue, hallway, handcraft, janitor, leak, mask, pedestrian, pipeline, plug, scrub, sewage, sewer, spark, sparkle, stumble, tangle, tank, technician, tunnel, underground, vacuum, zone Verb: drain, endanger, partition, pollute, plugging, stumbling, vacuuming, plug, stomp, stumble, unnoticed Adj.: civilian, delicious, electronic, polluted, proactive, fragrant, tasty, unorganized
Step 3 Words that have a meaning that is closely related to the situation and could be counted as technical terms in particular fields.	Noun: cable, cancer, chainsaw, circuit, coma, cord, diarrhea, drill, driller, fracture, glove, goggle, hardware, hazard, helmet, infection, lab, laboratory, outlet, passageway, scaffold, socket, toxin, virus, bacteria Verb: beware, disable, evaporate, hazard, infect, plug, unplug, contaminated, drilling, contaminating Adj.: contaminated, fatal, hazardous, radioactive, toxic
Step 4 Words that have very narrow meaning and used in limited context.	None of words falls in this category.

Lexical density

This section displays percentages of the lexical density of the entire text and of each vocabulary level produced in the speaking task by the students. Table 4 demonstrated the percentage of lexical density in one-minute speaking. There were 4,047 content words of 9,256 tokens, and the value of density was 43.72%. Among the content words, nouns (20.18%) were expressed the most frequently, followed by verbs (13.87%) and adverbs (5.52%). Whilst, adjectives (4.14%) were produced the least. These results suggested that the number of lexical nouns and verbs play an important role in the percentages of lexical density. This is unsurprising since these two parts of speech are fundamental units of sentence formation. In addition, to perform the one-minute speaking task, the students might have manipulated nouns and verbs to indicate the objects, actions, and place in the scene.

TABLE 4
Percentages of Lexical Density of the Entire Text

	Total tokens	Nouns	Verbs	Adjectives	Adverbs	Total content words
Numbers of tokens (N = 100)	9,256	1,868	1,284	384	511	4,047
Percentages of lexical density		20.18%	13.87%	4.14%	5.52%	43.72%

To dig deeper, investigating lexical density from the perspective of sophistication may enhance perceptive insights of vocabulary use. Table 5 presents percentages of density according to vocabulary levels. It is noticeable that noun density (10.00%) and verb density (11.05%) of GSL1000 were slightly different. The density of AWL (2.50%) was the lowest. At GSL 2000, the results provided obvious differences between noun density (3.72%) and verb density (1.78%). Considering the content words presented among the four levels, it is notable that they (nouns, verbs, adjectives, and adverbs) are highly dense especially at GSL1000, where the densities are 10.00%, 11.05%, 2.53% and 5.29% respectively. Despite nouns being the densest part of speech among the content words, its density appeared sparse in the higher-level lists. Lexical density of each type of content words of GSL2000 was also higher than those of AWL. To be exact, the density of the content words in GSL2000 are 3.72%, 1.78%, 0.70% and 0.19%; while, AWLs are 1.86%, 0.44%, 0.18% and 0.01% respectively.

TABLE 5
Percentages of lexical density of each vocabulary level

Density/Levels	GSL 1st 1000	Density (%)	GSL 2nd 1000	Density (%)	AWL	Density (%)	non-based list words	Density (%)
Nouns	926	10.00%	344	3.72%	172	1.86%	427	4.61%
Verbs	1023	11.05%	165	1.78%	41	0.44%	55	0.60%
Adjectives	243	2.53%	65	0.70%	17	0.18%	59	0.64%
Adverbs	490	5.29%	18	0.19%	1	0.01%	2	0.02%
Total content words	2682	28.98%	592	5.72%	231	2.50%	543	5.87%

In accordance with the results, lexical density indicated that EFL science and technology students' lexical banks, which the students can autonomously articulate in one-minute, belong to 1,000 word families. Overall, the density of nouns was densest (Table 4). However, the density of verbs was the highest at GSL1000. This might be the results of repeating the same verb while speaking. Verbs connect and complete the sentences for describing the situation. Even though densities of adjective and adverb were not high. These content words play essential roles of supplementing and completing the meaning of the speech.

Table 6 denotes the first 10 most frequent families in the speaking task. The most frequent word families in GSL1000 are *picture*, *sign*, *try*, *danger*, *man*, *see*, *wear*, *fall*, *eat*, and *head*. Words from GSL 2000 includes *warn*, *wire*, *clean*, *electric*, *brick*, *trip, hat*, *accident*, *police* and *smell*. AWL words show the task is oriented to technical contexts. The frequent families include *chemical*, *area*, *construct*, *equip*, *injure*, *label*, *site*, *confine*, *aware*, and *grade*. In the non-based list, the 10 most frequent words are *donut*, *helmet*, *guy*, *sewer*, *toxic*, *syrup*, *pancake*, *Homer*, *drain* and *hazard*. Among these, it is obvious that not all words are technical. For instance, Homer is a proper noun. Although, *syrup* and *pancake* were mentioned many times due to their obvious objects in the pictures, they are not technical words. In addition, *guy* is generally used in the spoken register and the word itself is not technical.

TABLE 6
10 Most Frequent Content Words (Families) Presented in Each Vocabulary Level

	GSL 1st 1000		GSL 2nd 1000		AWL 570		Non-Based List	
	words	Frequency	words	Frequency	Words	Frequency	Words	Frequency
1	picture	73	warn	43	chemical	40	donut	48
2	sign	63	wire	41	area	35	helmet	28
3	try	58	clean	40	construct	26	guy	27
4	danger	55	electric	36	equip	25	sewer	26
5	man	53	brick	31	injure	25	toxic	25
6	see	49	trip	29	label	18	syrup	24
7	wear	47	hat	23	site	18	pancake	19
8	fall	45	accident	20	confine	5	Homer	16
9	eat	44	police	18	aware	3	drain	14
10	head	42	smell	18	grade	3	hazard	14

Lexical diversity

This section shows the lexical diversity in the overall speech production of science and technology students. Figure 1 shows a comparison of the mean segmental TTR values (blue line) for samples of 35-50 words and values generated by the best-fitting value of D (yellow line) in Malvern and Richards's theoretical model. Theoretically, value of D can vary between 1 and 120. Low values of D indicate that the source text contains a lot of repetition, and is not lexically rich. High values of D indicate that the source text is lexically rich and tends not to repeat the same words over and over again (Meara, & Miralpeix, 2016 p. 33). The error statistic tells how close the data matches the model, and this figure should not be bigger than 0.01.

From the result, the value of D is 58.8. This means lexical diversity of the EFL is moderate. It implied that the students were likely to use repetition of words in their speaking. However, their lexical diversity was not very low; they were able to vary vocabulary in order to describe the situations in the picture. The error figure of 0.002 indicates that the actual data matched the model data and the program is appropriate for data analysis.

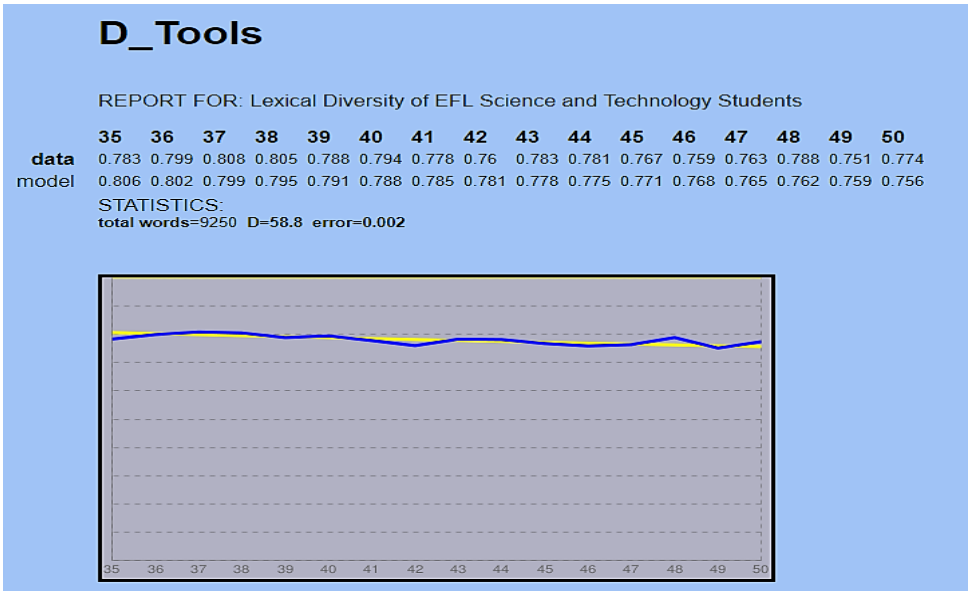


Figure 1. Lexical diversity in the speaking task of EFL science and technology students.

Discussion and Conclusion

This study investigated productive vocabulary of EFL science and technology students when performing one-minute speaking task. It was found that measuring different perspectives of the lexical richness (Read, 2000) manifested the learners' vocabulary proficiency. This includes ability to maximize and avoid reusing vocabulary, ability to use vocabulary in precise and sophisticated manners, and ability to provide necessary information about the scene. The following sections provide discussion of each aspect of the students' lexical richness.

Ability to maximize vocabulary, we found that EFL science and technology students' vocabulary diversity is moderately diverse. To elaborate, the students were capable of producing certain numbers of vocabulary in order to describe the situations in the picture. However, they were likely to use repetition of words in their speaking. The results implied that the students were not well familiar with the topic. In line with Yu (2009) and Skehan (2009), familiarity of topics as well as different aspects of speaking influence on achieving high lexical diversity. To a certain extent, the students experienced topics about safety in the classroom. However, the speaking task provided for them is a real-word task, which requires the extension of vocabulary knowledge to appropriately communicate in this particular context.

In this study, the overall lexical density of the speech production was higher than Ure's (1971) statement. Lexical density denoted the students' ability to deliver enough information in the pictures. Among the content words, the presence of nouns is densest, followed by verb, adverb, and adjective densities. The noun density was highest probably due to its functions as a subject, an object and a subject complement of a sentence. Another reason is the influences of the guidelines that emphasized on person, place, and objects presented in the pictures. Considering each vocabulary level, however, a proportion of lexical density showed lavish GSL1000 vocabulary especially verb density. The students took advantages of nouns and verbs to indicate materials or objects and actions in general to perform one-minute speaking. On the contrary, it is interesting that the students were able to produce limited words from GLS 2000, AWL and technical vocabulary. This might be the results of vocabulary orientation and quantity of target words rendered in different scenes, which is the limitation of this study.

We found that the students' spoken texts were chains of different levels of vocabulary. That is, the students exploited general vocabulary such as *sign*, *danger*, *man*, and *head* to answer questions of

what/who was in the picture and where the scene was. While, less general words such as *wire*, *brick*, *trip*, and *warn* were employed for delivering the details of the situations, specific objects or devices. They endeavored to express precise-meaning words in the contexts by drawing on academic words such as *chemical*, *confine*, and *injure*. Even though a small proportion of academic words presented, they helped convey more explicit and precise meanings closely related to the situations. Their meanings embraced the topic and made the speaking meaningful (Henrichs and Schoonen, 2009). The variety use of vocabulary from this study has shown the lexical sophistication to some extent.

Applying Chung and Nation's scales (2003) helped identify technical vocabulary in EFL speech productions. The results unveiled none highly technical vocabulary in the students' speaking. This is because the topic of the task is oriented to general technical contexts rather than highly technical contexts. Nonetheless, we found that students somewhat employed technical words such as *evaporate*, *hazard*, *infect*, *plug*, and *fatal* to strengthen relationship of words and the topic of the task. These words served the essence or main messages in the scenes and made the content completed. Thus, we could argue that the degree of vocabulary sophistication plays essential roles of clarity and precision of spoken products.

In conclusion, this study has presented students' speaking lexical richness through several measurements, and the data from each tool has been used to support each other. The lesson learned from the study may shed light for students and teachers. It suggests that students in ESP classroom contexts need significant awareness that they belong to a particular group. Additionally, they engage discipline knowledge that is closely tied to the specialized language use. People inside areas of language use of technical or professional spheres are anticipated understanding and exhibiting this language fluently (Coxhead, 2018). Therefore, they need to understand in details and more precision of vocabulary and language use. They need efforts to crave the knowledge that is imperative to attain academic, professional, or occupational goals.

The Authors

Thiwaporn Thawarom is a PhD candidate in Applied Linguistics, School of Liberal Arts, King Mongkut's University of Technology Thonburi. Her research interests include vocabulary learning and assessing, Content and Language Integrated Learning (CLIL) and learning strategies.

School of Liberal Arts
King Mongkut's University of Technology Thonburi (KMUTT)
Email: thiwaporn.t@mail.kmutt.ac.th

Wareesiri Singhasiri is an associate professor at the Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi. Her research interests cover learning strategies, learning styles, and research methodology.

School of Liberal Arts
King Mongkut's University of Technology Thonburi (KMUTT)
Email: wareesiri.sin@mail.kmutt.ac.th

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied linguistics*, *24*(4), 425-438.

Akbarian, I. H. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, *38*(3), 391-401.

Anthony, L. (2014). AntWordProfiler (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Anthony, L. (2015). TagAnt (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Benesch, S. (2001). *Critical English for academic purposes: Theory, politics, and practice*. Routledge.

Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice-Hall.

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language, 15*(2), 103-116.

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

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

Coxhead, A. (2013). Vocabulary and ESP. In. B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 115-132). West Sussex: John Willey & Sons Inc.

Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of Plumbing. *English for Specific Purposes, 51*, 84-97.

Crawford Camiciottoli, B. (2007). *The language of business studies lectures: A corpus-assisted analysis*. Amsterdam: John Benjamin.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics, 24*(2), 197-222.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes, 33*, 66-76.

Didau, D. (2013). *Black space: Improving writing by increasing lexical density*. Retrieved August 7, 2018, from https://learningspy.co.uk/writing/black-space-increasing-lexical-density/

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge, UK: Cambridge University Press.

Douglas, D. (2001). Language for specific purposes assessment criteria: Where do they come from? *Language Testing, 18*(2), 171-185.

Fiorito, L. (2005). *Teaching English for specific purposes (ESP)*. Retrieved June 18, 2018, from http://www.UsingEnglish.com.

Flowerdew, J., & Peacock, M. (2001). Issues in EAP: A preliminary Perspective. In M. H. Long & J. C. Richards (Eds.), *Research perspectives on English for academic purposes*. (pp. 8-24). Cambridge, Cambridge University Press.

Foster, P. (2009). Lexical diversity and native-like selection: The bonus of studying aboard. In. B. Richards, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application.* (pp. 91-106). Hampshire: Palgrave McMillan.

Goh, C. C. M. (2007). *Teaching speaking in the language classroom. Singapore*: SEAMEO Regianl Language Centre.

Guiraud, H. (1954). *Les Caractères Statistiques du Vocabulaire*. Paris: Presses Universitaires de France.

Halliday, M. A. K. (1985). *Spoken and written language*. Geelong Vict.: Deakin University.

Henrichs, L., & Schoonen, R. (2009). Lexical features of parental academic language input: The effect on vocabulary growth in monolingual Dutch children. In. B. Richards, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application*. (pp. 1-22). Hampshire: Palgrave McMillan.

Herdan, G. (1960). *Quantitative linguistics*. London: Butterworth.

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. *English for Specific Purposes, 33*, 54-65.

Huchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. NY: Cambridge.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Jin, N. Y., Ling, L. Y., Tong, C. S., Sahiddan, N., Philip, A., Azmi, N. H. N., & Tarmizi, M. A. A. (2013) Development of the engineering technology word list for vocational schools in Malaysia. *International Education Research, 1*(1), 43-59.

Johansson, V. (2008), Lexical diversity and lexical density in speech and writing: A developmental perspective, *Working Papers 53*, 61-79.

Kwary, D. A. (2011). A hybrid method for determining technical vocabulary. *System, 39*(2), 175-185.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757-786.

Laufer, B. 1989. 'What percentage of lexis is essential for comprehension'. In C. Lauren & M. Nordman (Eds.), *Special language: From human thinking to thinking machines* (pp. 316-323). Clevedon: Multilingual Matters.

Laufer, B., & Nation., P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*, 307-322.

Li, J., & Wai Chun, C. K. (2012). Assessing the roles of metacognitive knowledge and vocabulary size in EFL reading comprehension. *The Journal of Asia TEFL, 9*(4), 1-26.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190-208.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (Eds.), *Evolving models of language*. (pp. 58-71) Clevedon: Multilingual Matters.

Malvern, D. D., & Richards, B., J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language testing, 19*(1), 85-104.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Houndmills, NH: Palgrave Macmillan.

McCarthy, P.M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)* (Unpublished doctoral dissertation). University of Memphis. Retrieved from https://umdrive.memphis.edu/pmmccrth/public/Phil's% 20papers.htm?uniq-xq6brv

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods, 42*(2), 381-392.

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Clevedon: Multilingual Matters.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes, 25*(2), 235-256.

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

Nation, P. (2008). *Teaching vocabulary: Strategies and techniques*. Boston: Heinle Cengage.

Nation, P. (2013). *Learning vocabulary in another language*. (2nd ed.). Cambridge: Cambridge University Press.

Paltrideg, B., & Starfields, S. (2013). *The handbook of English for specific purpose*. Boston: Wiley-Blackwell.

Paulus, T., Lester, N. J., & Dempster, P. (2014). *Digital tools for qualitative research*. SAGE Publications Ltd.

Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOl Quarterly*, 77-89.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan.

To, V., Fan, S., & Thomas, D. (2013). Lexical density and readability: A case study of English textbooks. *Internet Journal of Language, Culture and Society, 37*, 61-71.

Torruella, J., & Capsada, R. (2013). Lexical statistics and typological structures: A measure of lexical richness. *Procedia-Social and Behavioral Sciences, 95*, 447-454.

Ure, J. (1971), Lexical density and register differentiation. In G. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443-452). London: Cambridge University Press.

Van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Veenstra, J., & Sato, Y., (2018). Creating an institution-specific science and engineering academic word list for university students. *The Journal of Asia TEFL, 1*(15), 149-162. doi: http://dx.doi.org/10.18823/asiatefl.2018.15.1.10.148

Ward, J. (2009) A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes 28*, 170-182

Watson Todd, R. (2017). An opaque engineering word list: Which words should a teacher focus on? *English for Specific Purposes, 45*, 31-39.

West, M. (1953) *A general service list of English words*. London: Longman.

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied linguistics*, *31*(2), 236-259.

Zhang, Y. (2014). A corpus-based analysis of lexical richness of Beijing Mandarin speakers: Variable identification and model construction. *Language Sciences, 44*, 60-69.

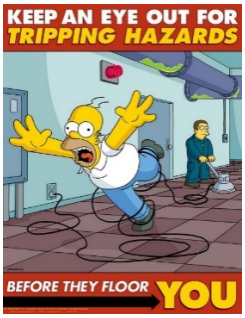
Zipagan, M, N., & Lee, K, R., (2018). Korean English learners' use of lexical bundles in speaking. *The Journal of Asia TEFL, 15*(2), 276-291.

Appendix

Pictures for the Speaking Task

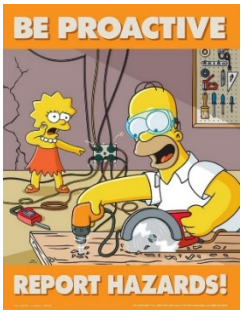
Picture 1

Picture 2

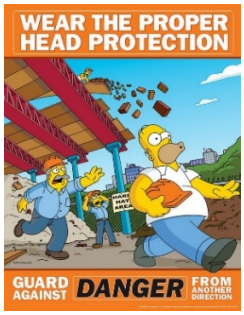


Picture 3

Picture 4



Picture 5



Picture 1, 4, and 5 were Retrieved from https://piximus.net/others/the-simpsons-are-here-to-teach-you-about-work-safety
Picture 2 retrieved from https://oshmatters.files.wordpress.com
Picture 3 retrieved from
https://simpsonswiki.com/w/index.php?title=The_Simpsons_Safety_Posters&printable=yes