Lexical Density and Readability:
A Case Study of English Textbooks

Vinh To

University of Tasmania

Si Fan

University of Tasmania

Damon Thomas

University of Tasmania

Abstract

This paper aimed to examine the lexical density and readability of four texts from English textbooks known as Active Skill for Reading (Anderson, 2003a, 2003b, 2003c, 2003d) at elementary, pre-intermediate, intermediate and upper-intermediate levels. The study applied three methods in determining lexical density and readability as proposed by Halliday (1985), Ure (1971) and Flesch (1948). The analysis revealed that three of the four reading texts were of high lexical density, apart from the text for upper-intermediate level. In terms of readability, the levels of texts corresponded to readability levels. However the highest level did not entail the topmost readability. There was little evidence of an increase of lexical density and readability in accordance with the increase of text levels as well as little indication relating to the connections between text levels, readability and lexical density. With reference to the methods employed, Halliday's measurement of lexical density - based on the clause - had a significant correlation with Ure's measurement, and a medium relation with Flesch's Reading Ease Scale, whereas Ure and Flesch's formulas showed no correlation.

Key words: *lexical density, readability, linguistic complexity, reading texts, English textbooks*

Introduction

"Language describes complex phenomena, but is also itself a complex phenomenon" (Hendrikse & Van Zweel, 2010, p. 409). Though linguistic complexity is fundamental to the current science, its precise definition has still been an open issue (Kwapien, 2010). In addition, while the notion of complexity is central to literacy and language education, there is no standard linguistic measure towards it (Rimmer, 2008). This study explores the lexical density of several texts, which is recognised as a significant factor of complexity of written language (Halliday, 1985b). Restricted by its limited scope, the study investigates four extracts from reading passages in four textbooks written for elementary, pre-intermediate, intermediate and upper-intermediate levels. The study's purposes are to identify how lexical density and readability are realised in the selected extracts and to investigate the relationships between lexical density, readability, and text levels. The study also looks at the correlation between measurements of lexical density and readability, and uses a combination of two methods in examining lexical density as indicated by Ure (1971) and Halliday (1985b). The results, afterward, are compared with Flesch's (1948) Reading Ease Scale , which is demonstrated as a reliable measure of readability by England, Thomas, and Paterson (1953).

Literature Review

The concept of *density* refers to a kind of complexity that results from the development of words. In other words, this relates to the notion of lexico-grammar in terms of the level of wording in language (Halliday, 1985).There are at present several different measurements of lexical density. Originally, it

was proposed by Ure (1971) that lexical density should be treated as the proportion of the number of lexical items per the number of running words. This formula was refined by Halliday (1985) as his first approximation to measure lexical density, and was then further developed by O'Loughlin (1995). The following example provides a brief illustration of how the concept of lexical density has developed.

E.g. *The **Trust** has **offered advice** to **local government authorities** on **cemetery conservation*** (Halliday, 1985, p.61).

There are eight lexical items and four grammatical items in this sentence, giving a proportion of eight lexical items out of twelve items in total, and using Ure's (1971) original method, the lexical density would be sixty-seven per cent or 0.67.

However, Halliday (1985), the originator of Systemic Functional Linguistics, proposes another formula in determining lexical density of texts based on clauses. He states that "lexical density is the number of lexical items as a ratio of the number of clauses" (Halliday, 1985, 67). For instance:

E.g. /*The **basic 'stuff'** of **living organisms** is **protoplasm***/. *There is no **set composition** of this /and it **varies** between one **individual** and the next/* (Halliday, 1985, p.67).

There are nine lexical items and three clauses in this text, giving the ratio of nine out of three, which shows the lexical density of the text to be 3.0.

In order to measure lexical density, it is necessary to distinguish grammatical items from lexical items and the differences between them. Grammatical items, or 'function words', come in a closed system comprising of determiners such as articles, pronouns, most prepositions, conjunctions, some classes of adverb and finite verbs (Cindy & James, 2007; Halliday, 1985b). In contrast, lexical items, or 'content words', are named *lexical* as they function in lexical sets, that is to say, they are part of an open system rather than closed sets (Halliday, 1985). Traditionally, nouns, verbs, adjectives and adverbs are the four word classes belonging to lexical items since they have autonomous meaning even in isolation and new members can be added to these categories (T.Le, Yue, & Le, 2011). It is interesting to note that Halliday (1985) uses the term 'items' rather than 'words' when discussing lexical items and grammatical items, since he argues that they may contain more than one word in the usual sense. Taking several phrasal verbs such as *stand up, take over,* or *call off*, each of these consists of two words, a lexical verb and a preposition, but Halliday treats them as a lexical item. This is quite a contrast to the views of Ure (1971), who counts them as two separate words, one being the lexical word *stand, take, or call,* and the other being the preposition *up, over,* or *off* respectively. Halliday (1985) makes clear that in most cases it is unambiguous to define which the lexical item or grammatical item is, however, in some cases it is more difficult to define, English prepositions and certain classes of adverb, *always* and *perhaps* for instance, which fall on this borderline.

Regarding previous research on lexical density, various texts have been intensively examined and different outcomes have been achieved. With reference to lexical density in written texts, Narelle et al. (1994) investigated difficulties encountered in students' statistics examinations in their first year at an Australian university. They analysed 186 students' examination papers with the hypothesis that language difficulty would influence student performances in the examinations. Then, the lexical density of the texts was determined and compared with the performances of students obtaining full points and those who obtained no points. However, the finding was unexpected. There was no connection between student achievement and the linguistic complexity of the questions in terms of lexical density. Likewise, Cheryl (1995) examined the correlation of lexical proficiency to the quality of ESL compositions written by students at Indiana University in the matters of lexical density, lexical errors and lexical variations. The essay topics given to students had a relatively average lexical density. As four lexical measurements were used in the study, lexical density could be the most concisely interpreted. The results indicated a non-significant relationship between the proportion of content words and the quality of essays. It should be added that Vidakovic and Barker (2009) found a similar result. The aims of their research were to explore the lexical progression in general purpose English target texts. They investigated a dataset of 100 passing Cambridge Certificates in ESOL Skills for Life with writing scripts of 36 million words taken from the Cambridge Learner Corpus. Thenceforward, 20 students from each of five levels (Entry 1, Entry 2, Entry 3, Level 1, Level 2) were selected to represent levels A1-C1 of the Common European Framework of Reference. Students wrote between 53 and 1003 words which provided 2,300–14,000 words per level. The results suggest that although the average length of words and sentences, the number of different words, as well as lexical variation increased with proficiency, lexical density did not differentiate according to proficiency levels.

Aside from written texts, lexical density has been scrutinised in spoken texts as well (Hilary, 2001; John & Paul, 2002). John and Paul (2002) surveyed the transcriptions of 88 recorded IELTS Speaking Tests from band 4 to band 8 at 21 test centres worldwide in order to access the candidates' vocabu-

lary use in terms of lexical density, lexical diversity, lexical sophistication, and the use of formulaic language. The results reveal a decreasing pattern from band 8 to band 4 regarding the lexical features indicated as well as a substantial disparity within bands which indicates that these lexical statistics do not provide a reliable foundation to assess speaking proficiency levels. It also shows that candidates at the advanced level used fluently a variety of formulaic expressions containing many high-frequency words rather than low-frequency words, while candidates at band 4 rarely used formulaic language (John & Paul, 2002). Hilary (2001) analysed the BASE corpus of authentic academic speech with 30 undergraduate lectures delivered at Warwick University between 1998 and1999 to examine lexical density and the relationship between lecturing purpose and delivery style. The findings show that lectures which were delivered more quickly tended to be sparser, while lectures produced more slowly tended to be denser. Moreover, the professional speakers delivered speeches with the highest densities and the language used in these speaking contexts was formulaic and pre-rehearsed.

It is worth noting that lexical density in written and spoken languages has been compared (Belinda, 2007; Halliday, 1985b; Yu, 2007). As noted earlier, there are several formulas to calculate lexical density in a text. It can be expressed either by the proportion of the number of content words per the total words (Ure, 1971; Halliday,1985; O'Loughlin, 1995) or as a ratio of the amount of lexical words per clause (Halliday, 1985). Irrespective of the parameters adopted, research over the past years has indicated that, overall, writing is much denser lexically compared to speech (Belinda, 2007; Halliday, 1985b; Yu, 2007). Interestingly, the lexical density of various languages other than English has also been examined, including Bantu languages (Oliver, 2007), Swedish and German (Linnarud & Thoursie, 2008) and Dutch (Henrichs, 2010).

Methodology

Aim and objectives:

This study aims to investigate the lexical density and readability of selected reading extracts from English textbooks used for teaching English to speakers of other languages. The three objectives are outlined as follows:

- Objective 1: To find out the levels of lexical density and readability among selected reading texts in English textbooks and the relation between them within these texts;
- Objective 2: To identify the changes of lexical density and readability across levels of the texts in English textbooks;
- Objective 3: To discover the correlation among the measurements of lexical density and readability.

Research questions

Equivalent to the three objectives outlined above, the three research questions are recognised:

- Question 1: How is lexical density and readability presented in the selected texts in English textbooks? And what is the relation between them?
- Question 2: Do the lexical density index and readability increase in accordance with the text levels?
- Question 3: Is there a strong correlation between the methods adopted to access the lexical density and readability?

Description of chosen methods

The study employs the two methods proposed by Ure (1971) and Halliday (1985) which are discussed above as central measurements to lexical density exploration in texts.

Formula 1 (Ure's method):

$$\text{Lexical density} = \frac{\text{number of lexical items x 100}}{\text{total numbers of words}}$$

(Halliday, 1985b; O'Loughlin, 1995; Ure, 1971)

Regarding this measurement, if the number surpasses forty per cent, it accounts for higher lexical density.

Formula 2 (Halliday's method):

Lexical density =	number of lexical items
	numbers of ranking clauses

(Halliday, 1985)

In terms of the second method, Halliday (1985) found that a typical average lexical density for written texts is between 3 and 6, depending on the text formality. The higher the index, the more difficult the text is. The figures for spoken language, conversely, fall to between 1.5 and 2.

The statistics obtained from the above methods are compared with Flesch's (1948) Reading Ease Scale which measures sentence length and the number of syllables per 100-word passages (Courtis & Hassan, 2002). Its concrete formula is stated below, in which *wl* indicates the total number of syllables of the words in the passage and *sl* indicates the average number of words per sentence.

Formula 3 (Flesch's method):

$$\text{Reading ease} = 206.835 - 0.846\,wl - 1.015\,sl$$

(Wimmer & Dominick, 2006, p. 344)

In this method, the number 42.98 indicates the middle of difficulty categories (Courtis & Hassan, 2002), while the other numbers are shown in the following table and compared with educational attainment levels in the American educational system.

Table 1. *Original Flesch reading ease* (Courtis & Hassan, 2002, p. 406)

Flesch reading ease	Description of style	Educational Attainment Level (US)
0-30	Very difficult	Postgraduate
30-50	Difficult	Undergraduate
50-60	Fairly difficult	Grade 10-12
60-70	Standard	Grade 8-9
70-80	Fairly easy	Grade 7
80-90	Easy	Grade 6
90-100	Very easy	Grade 5

Concerning the first two methods, the ascending numbers indicate texts with a higher level of difficulty, while with the third method, increasing statistics reveal lower levels of text difficulty. Since this paper aims to identify the correlation between the three methods, it has been deemed more convenient and justifiable to convert the Flesch Reading Ease to increase respectively with the level of text difficulty. As a consequence, the adapted Flesch Reading Ease employed in this study is described in the table below.

Table 2. *Flesch Reading Ease used in this study*

Flesch reading ease	Description of style	Educational Attainment Level (US)
0-10	Very easy	Grade 5
10-20	Easy	Grade 6
20-30	Fairly easy	Grade 7
30-40	Standard	Grade 8-9
40-50	Fairly difficult	Grade 10-12
50-70	Difficult	Undergraduate
70-100	Very difficult	Postgraduate

Software adopted

In this study, Microsoft Word and Excel were employed to deal with some counting and abstract statistics. First, Microsoft Word was used to calculate the total words in each text, and to show the readability statistics realised by Flesch's (1948) formula. Later, Microsoft Excel was employed to analyse the data and determine the correlation between measurements.

Description of the corpus

This small corpus consisted of four reading extracts taken from four English textbooks commonly used in teaching reading skills known as Active Skills for Reading (Anderson, 2003a, 2003b, 2003c, 2003d).The four books are set for use at the following four levels respectively, elementary, pre-intermediate, intermediate and upper-intermediate. Each chosen text was extracted from the first unit of each book and contain around 160-173 words, making a corpus of 660 words in total. With the purpose of fulfilling the effectiveness of data analysis, each text is entitled as follows.

Text 1 was taken from unit 1, book 1 (Elementary);

Text 2 was taken from unit 1, book 2 (Pre-intermediate);

Text 3 was taken from unit 1, book 3 (Intermediate);

Text 4 was taken from unit 1, book 4 (Upper-intermediate).

Data analysis

This research features the use of three formulas as proposed by Ure (1971), Halliday (1985) and Flesch (1948) to examine the lexical density of texts chosen with the purpose of achieving relevant and reliable results. The study, then, determines the correlations between these methods to indicate how the statistics work together. The procedure is described in the following steps. First, lexical items were manually counted based on Halliday's rules of determining lexical items. At the same time, the total words were calculated using Microsoft Word. Ranking clauses were determined regarding finite verbs. Then, Halliday (1985) and Ure's (1971) formulas were applied to obtain the lexical density index which was later compared with Flesch (date) Reading Ease Scale with the use of Microsoft Excel.

Findings and discussion

Findings

Lexical density and readability among chosen reading texts from English textbooks

With reference to the three lexical measures adopted in this study, lexical density index can be interpreted concisely. The data analysis shows that the selected texts were of high lexical density with the exception of text 4. Initially, using Ure's formula it is shown in the table below that the lexical index of the four texts ranged from the lowest 45.5% for Text 4 to the highest 53.8% for Text 2, which was above the average number 40% as proposed by Ure (1971). Regarding Halliday's (1985) method, the statistics demonstrate the typical density in written language which ranged from 3 (Text 4) to 4.8 (Text 2). Finally, the Flesch Reading Ease Scale showed that Text 1 was fairly easy for students to read while Texts 3 and 4 were relatively difficult, and Text 2 was the most challenging. It is significant to note that the three methods were identical in describing Text 2 (pre-intermediate) as the most difficult and the densest among the dataset. Despite this, there are several considerable disparities in the figures interpreted by these formulas. Firstly, Text 1 (elementary level) is the easiest reading passage according to Flesch Reading Ease; however it also has a high percentage of density (53.2%) when considered with Ure's (1971) method. Secondly, Halliday (1985) and Ure's (1971) measurements indicate that Text 4 (upper-intermediate level) possesses the lowest lexical density among the four texts, which is quite contradictory to Flesh's (date) method considering Text 1 (elementary level) is apparently the simplest one. Finally, in the light of Educational Attainment Level (US), Text 2 is recommended as suitable for undergraduate education, Texts 3 and 4 are appropriate for students at the level of grades 10-12, while Text 1 is appropriate for students of grade 7. The results are summarised in the following table.

Table 3. *Lexical density among the four texts*

Texts		1 (Elementary)	2 (Pre-intermediate)	3 (Intermediate)	4 (Upper-intermediate)
Total words		173	160	162	165
Lexical words		92	86	75	75
Functional words		81	74	87	85
Ranking clauses		26	18	20	25
Lexical density	Ure's method (%)	53.2	53.8	46.3	45.5
	Halliday's method	3.5	4.8	3.8	3.0
	Flesch's method	25.2	53.3	48.1	43.9

The changes of lexical density and readability across levels of the texts in English textbooks

Considering the range of lexical density across levels of the reading extracts, Ure (1971), Halliday (1985) and Flesch's (1948) methods featured relatively similar patterns. As can be seen in the three figures below, the lexical density increased from Text 1 to Text 2 before decreasing in Text 3 and further decreased in Text 4. While the first method was slightly different, the latter methods were relatively similar to each other with respect to the overall trend. Concerning Ure's (1971) measurement, the lexical density increased slightly from 53.2% in Text 1 to 53.8% in Text 2, then decreased to 46.3% in Text 3 and 45.5% in Text 4. For Halliday (1985) and Flesch's (1948) formulas, the figures began at the lower points (3.5 and 25.2 respectively), before increasingly considerably to peaks of 4.8 and 53.3 respectively, followed by a gradual decrease to 3 and 43.9 as the final points. Although it may be assumed that lexical density changes in accordance with the set difficulty levels of texts, the results of this study do not support such a notion. Specifically, Text 3 and Text 4, which represent the higher levels of intermediate and upper-intermediate learners, were less dense lexically when compared with Text 2 (Pre-intermediate level). Text 4 (upper-intermediate) was the least dense overall, even lower in lexical density than Text 1 (elementary), as measured by Ure (1971) and Halliday's (1985) formulas. Overall, it is concluded that higher levels of texts do not necessarily guarantee higher lexical density indexes, that is to say, the levels of texts analysed in this study did not match the expected indexes of lexical density. The levels of lexical density for the four texts are illustrated in the three figures below.
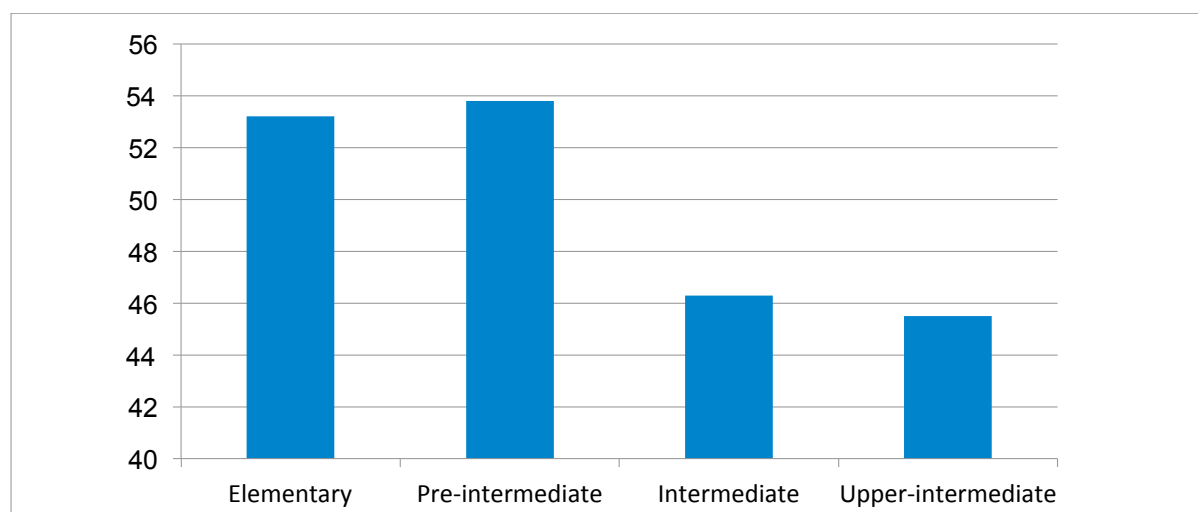

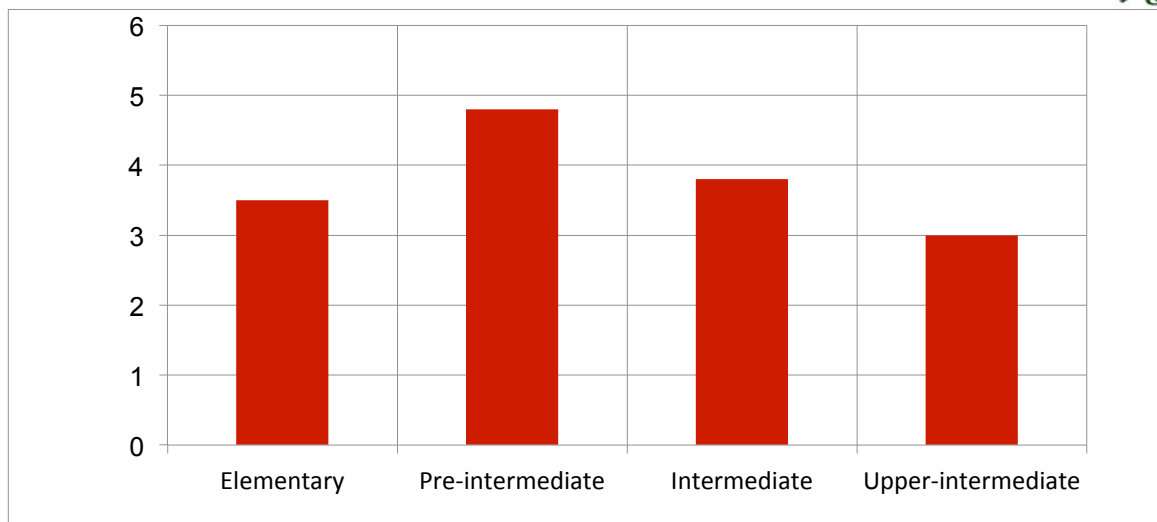*Figure 1.* Lexical density, Ure's method (%)
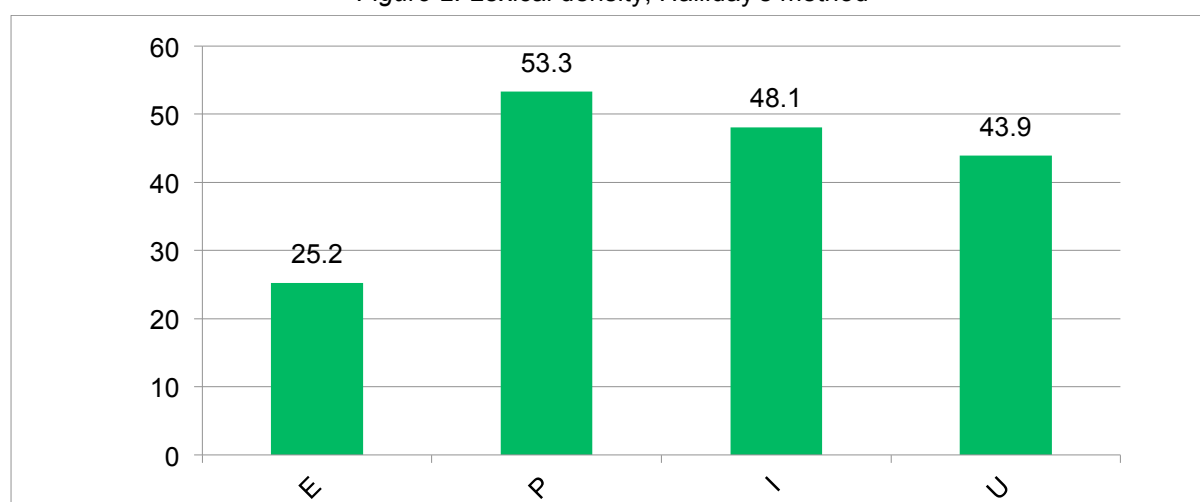
Figure 2. Lexical density, Halliday's method

Figure 3. Flesh Reading Ease Scale

The correlations among the measurements of lexical density and readability

With the purpose of obtaining the most accurate findings, this study employed three various methods as mentioned above to analyse the data. It is essential to notice the correlation between these methods to more comprehensively understand how the data works. As can be seen in Table 4 below, there is a strong correlation between Halliday (1985) and Ure's (1971) formulas, with a correlation coefficient of 0.6. Both methods revealed that Text 4, which is recommended for advanced English learning, had the lowest lexical density index (3.0 and 45.5 respectively), whilst Text 2, presumed suitable for learners at pre-intermediate level, was recognised as the most dense of the four texts. Additionally, Halliday's (1985) method and Flesch Reading Ease featured a medium correlation coefficient of 0.5. As can be observed in the figures above, these two methods share a similar pattern although they are somewhat different. Particularly, Flesch Reading Ease Scale revealed Text 1 (elementary) as the easiest of the four texts, while Halliday's (1985) method showed Text 4 (upper-intermediate) to have the lowest lexical density. It is interesting to note that there was no correlation between Flesch (1948) and Ure's (1971) methods, which stated a negative correlation coefficient of minus 0.3.

Table 4. *Correlation coefficient between the methods*

Texts	Lexical density			Correlation coefficient		
	Ure's method	Halliday's method	Flesch's method	Halliday & Ure's method	Halliday & Flesch's method	Flesh & Ure's method
1	53.2	3.5	25.2			

2	53.8	4.8	53.3	0.6	0.5	-0.3
3	46.3	3.8	48.1			
4	45.5	3.0	43.9			

Discussion

Question 1: How is lexical density and readability presented in the selected texts in English text-books? What is the relation between lexical density and readability within these texts?

With a careful analysis of lexical density in the four reading extracts, the results reveal that almost all of the texts featured a high lexical density index apart from Text 4. As far as Table 3 is concerned, the dominant feature interpreted by the three measurements is that Text 2 (pre-intermediate) featured the topmost position in terms of lexical index. The results of this study suggest this text is actually more appropriate for native English speakers at the undergraduate level, when assessed with regards to the American educational attainment levels. Likewise Text 3 has a high lexical density when considered with Halliday (1985) and Ure's (1971) measurements, and is rated as rather difficult in terms of Flesch's Reading Ease Scale. As for Text 1 and Text 3, the figures are fairly diverse. Text 1 represents the easiest reading extract by Flesch's (1948) method, and is fairly simple by Halliday's (1971) method. It might be suitable for children reading at the Grade 5 level, regarding educational attainment levels in the United States, whereas it took the second ranking of high lexical density texts when using Ure's (1971) formula. As can be seen on Table 3, Text 4 possessed the lowest lexical density with reference to Ure (1971) and Halliday's (1985) methods, but it is still relatively difficult in terms of readability and is arguably suitable for use by high school students. To conclude, a text in an English textbook with high lexical density can be easy to read, while in some cases, texts with a low density can be difficult to read and vice versus. Below is the summary of the relation between lexical density and readability of the four texts.

Table 5. *The relation between lexical density and readability*

Text	Lexical density	Readability
1	high	easy
2	highest	most difficult
3	high	fairly difficult
4	low	fairly difficult

Question 2: Does the lexical density index and readability change gradually in accordance with the text levels?

Firstly, pertaining to lexical density, the numbers from Table 3 above show that the texts at elementary, pre-intermediate and intermediate levels were lexically dense, while the upper-intermediate text featured a low lexical density index. Therefore, it can be concluded that lexical density in texts from English textbooks does not always match its level. A lower-level text might be of high lexical density, whereas a low density text might be recommended for English learning at higher levels. Secondly, concerning readability, the statistics reveal another interesting result. The text at pre-intermediate level was the most demanding in comprehension in comparison with intermediate and upper-intermediate levels. That is to say, the readability did not become more difficult with the increase of the text levels. In general, lexical density and readability within the corpus of this study did not change in accordance with the increase of text levels. The discussion is clarified in the table below as follows:

Table 6*. The relation between lexical density, readability and text levels*

Level	Lexical density	Readability
Elementary	high	easy
Pre-intermediate	highest	most difficult
Intermediate	high	fairly difficult
Upper-Intermediate	low	fairly difficult

Question 3: Is there a strong correlation between the methods adopted to access the lexical density and readability?

What is inferred from the results in Table 4 is that there was a moderately strong correlation between Halliday (1985) and Ure's (1971) measurements of lexical density. In this study, these methods shared similar trends in representing the lexical densities featured in the four investigated texts. In addition, Flesch Reading Ease featured an average correlation with Halliday's (1985) method. This corresponded to a similar pattern in lexical density change in addition to some contradiction in lexical density index and readability. An example for this can be seen in Text 1 and Text 4 in Table 4. Finally, when considering Ure's (1971) lexical density formula and Flesch Reading Ease, there was no indication found of any significant connection.

Conclusion

This paper performed an investigation into the lexical density (Halliday, 1985; Ure, 1971) and readability (Flesch, 1948) of four reading extracts from four English textbooks (Anderson, 2003a, 2003b, 2003c, 2003d). The results indicate that this corpus was lexically dense apart from Text 4, which is set for the upper-intermediate level. As for readability, three texts including Text 2, Text 3 and Text 4 were syntactically difficult to read, while the simplest was Text 1 (elementary). It is also stated that Text 2 (pre-intermediate) featured the highest lexical density and was the most demanding text in terms of readability. Moreover, the study reveals an absence of convincing proof about any strong link between lexical density and readability, between lexical density and English reading comprehension text levels, as well as these text levels and their overall readability. As a final point, correlations between the formulas utilised were explained, with findings indicating that Halliday's (1985) method had a considerable connection with Ure's (1971) method, and a medium connection with Flesch Reading Ease, whilst the latter measurements did not correlate in any significant way. These findings suggest that, contrary to the common sense that lexical density, the typical kind of complexity in written language (Halliday, 1985), may increase steadily with the text levels and their readability, there was no strong relation between them in the reading comprehension texts in this study. It is argued that the linguistic complexity of texts in English textbooks may depend on other grammatical aspects rather than lexical density. Therefore, it is worth pointing out that further studies should be carried out to investigate the complexity of language in English textbooks so as to provide educators an effective method to choose appropriate texts in English teaching or English curriculum design.

References

Anderson, N. J. (2003a). *Active Skills for Reading, Book 1*. Boston: Heinle Cengage.

Anderson, N. J. (2003b). *Active Skills for Reading, Book 2*. Boston: Heinle Cengage.

Anderson, N. J. (2003c). *Active Skills for Reading, Book 3*. Boston: Heinle Cengage.

Anderson, N. J. (2003d). *Active Skills for Reading, Book 4*. Boston: Heinle Cengage.

Belinda, C. C. (2007). *The Language of Business Studies Lectures: A Corpus-Assisted Analysis*. The Netherlands: John Benjamins Publishing Company.

Cheryl, A. E. (1995). The Relationship of Lexical Proficiency to the Quality of ESL Compositions. *Journal of Second Language Writing, 4* (2), 139-155.

Cindy, C., & James, P. (2007). *The Psycholigical Functions of Function Words*. New York: Psychology Press.

Courtis, J. K., & Hassan, S. (2002). Reading Ease of Bilingual Annual Reports. *Journal of Business Communication 39*(4), 394 - 413.

England, G. W., Thomas, M., & Paterson, D. G. (1953). Reliability of the original and the simplified flesch reading ease formulas. *Journal of Applied Psychology, 37*(2), 111-113. doi: 10.1037/h0055346

Halliday, M. A. K. (1985b). *Spoken and written language*. [Waurn Ponds], Vic: Deakin University.

Hendrikse, R., & Van Zweel, H. (2010). A phylogenetic and cognitive perspective on linguistic complexity. *Southern African Linguistics & Applied Language Studies, 28*(4), 409-422. doi: 10.2989/16073614.2010.548017

Henrichs, L. F. (2010). Academic language in early childhood interactions : a longitudinal study of 3-to 6-year-old Dutch monolingual children *Academic language in early childhood interactions* (pp. 45-86): The University of Amsterdam.

Hilary, N. (2001). A corpus-based analysis of academic lectures across disciplines. In J. a. I. A. Cotterill (Ed.), *Language Across Boundaries* (pp. 201-218). London: BAAL in association with Continuum Press.

John, R., & Paul, N. (2002). An investigation of the lexical dimension of the IELTS Speaking Test. *IELTS Research Reports, Vol.6*.

Kwapien, J. J., Drozdz, S. S., & Orczyk, A. A. (2010). Linguistic Complexity: English vs. Polish, Text vs. Corpus. *Acta Physica Polonica, A, , 117*(4), 716-720.

Linnarud, M., & Thoursie, S. (2008). English and German in Swedish Classrooms: Writing in the two languages compared. *Nordic Journal of English Studies, 7*(2), 75-98.

Narelle, F. S., Leigh, N. W., Roslyn, K. G., & Gillian, P. (1994). Analysis of Student Performance in Statistics. Sydney: University of Technology.

O'Loughlin, K. (1995). Lexical density in Candidate output on two versions of An oral Proficiency Test. *Melbourne Papers in Language Teaching*, 26-48.

Oliver, S. (2007). Lexical Density in Oral versus Written Rangi Texts. *SOAS Working Papers in Linguistics, 15*, 173-184.

Rimmer, W. (2008). Putting grammatical complexity in context. *Literacy, 42*(1), 29-35. doi: 10.1111/j.1467-9345.2008.00478.x

T.Le, Yue, Y., & Le, Q. (2011). *Linguistic complexity and its relation to language and literacy education*. New York: Nova Science Publishers, Inc.

Ure, J. (1971). Lexical density and register differentiation. *Applications of linguistics*, 443-452.

Vidaković, I., & Barker, F. (2009). *Lexical development across second language proficiency levels: a corpus-informed study.* Paper presented at the Proceedings of The BAAL Annual Conference, Newcastle University.

Wimmer, R. D., & Dominick, J. R. (2006). *Mass media research : an introduction* (8th ed.). Belmont, Calif. ; London: Thomson/Wadsworth.

Yu, G. (2007). Lexical Diversity in MELAB Writing and Speaking Task Performances. *Spaan Fell* Retrieved 9, 2012

Appendix: Extract from the reading passage

Text 1: Extract from the reading passage "Food That Makes You Feel Good"

(Active Skills for Reading: Unit 1, Book 1, Elementary by Neil. J. Anderson, 2003, p.3)

Food is life; it gives us the nourishment we need to stay alive and be healthy. Usually, we eat because we are hungry or need energy. Brian Wansink, a professor at the University of Illinois, says we also eat certain foods because they make us feel good and remind us of happy memories. Wansink calls this kind of food *comfort* food. For some people, ice cream is a comfort food. For others, a bowl of noodle soup makes them feel good.

How does a food become comfort food? Professor Wansink believes that we connect food with important times, feelings, and people in our lives. "When I was a child, my mother made a delicious soup; I loved it. Now, I often eat this soup when I am tired or worried, and it helps me feel better," says one of Wansink's coworkers. Do men and women choose different comfort foods? Wansink's research at the University of Illinois says "yes". In his study, the favourite comfort food for both men and women was ice cream.

Text 2: Extract from the reading passage "The History of Chocolate"

(Active Skills for Reading: Unit 1, Book 2, Pre-intermediate by Neil. J. Anderson, 2003, p.3)

In the sixteenth century, chocolate was taken back to Spain by Hernando Cortez, another explorer. The Spanish people added ingredients such as sugar and vanilla to make it sweet. It later spread to France in the seventeenth century after the marriage of Louis XIII to the Spanish princess, Anna, who loved chocolate. The popularity of chocolate continued to spread further across Europe and the Americas. The only Asian country to adopt it at that time, though, was the Philippines, which the Spanish

invaded in the sixteenth century. As chocolate became more popular, the demand for people to work on the cocoa plantations increased. Slaves were brought to the Americas from Africa to farm the cocoa. Eventually, the cacao tree was taken to Africa and cultivation began there. Today, the African plantations provide almost seventy percent of the world's cacao, compared with one and a half precent from Mexico. Who would have thought that chocolate candy bars had so much history?

Text 3: Extract from the reading passage "Are You a Procrastinator?"

(Active Skills for Reading: Unit 1, Book 3, Intermediate by Neil. J. Anderson, 2003, p.7)

According to recent studies, there are three main reasons that students procrastinate. First, many have poor time-management skills and often try to do too much in too little time. In the end, these students often feel overwhelmed and will put off doing many things they need to. Another reason that students procrastinate is because they feel a subject is boring or because they have difficulty concentrating on an assignment. These students will often avoid doing something because they don't like it. A third reason that many students procrastinate is because they are very anxious about doing assignments well enough. These students often worry that their work will never be as good as it should be and fear failure of any kind. Unfortunately, trying to do everything perfectly can often cause these students to put off doing any work at all.

Do you recognize any of these signs in yourself? If so, you may want to do something about your tendency to procrastinate.

Text 4: Extract from the reading passage "Creating a Global Resume"

(Active Skills for Reading: Unit 1, Book 4, advanced by Neil. J. Anderson, 2003, p.3)

Most employers who want to hire from abroad must be able to certify to their local government that they are unable to find locals with the required skills necessary to do the job. The fastest way to be hired abroad is either to seek a country where there is a shortage of people with your skills or to be an "intra-company" transfer to another country. Be aware that obtaining a work permit can take a few months.

To be successful, and to enjoy your experience abroad, you must be flexible, open-minded, and both eager and willing to learn new ways of doing things. People everywhere appreciate individuals who are interested in getting to know them and learn about their ways of doing things. Enormous cultural faux pas are forgiven of individuals who are making honest attempts to fit in. Be patient and observant. Ask questions to demonstrate your interest in learning and broadening your horizons. Remember that you represent your country to everyone you meet.