A COMPARATIVE STUDY ON COMPLEXITY OF TEXTS IN COLLEGE ENGLISH TEXTBOOK AND IN CET 4

Fu Lei

National University of Uzbekistan Named after Mirzo Ulugbek, Uzbekistan
Inner Mongolia University of Finance and Economics, China

ABSTRACT
This study evaluates the text complexity of passages from Textbook 1 in comparison to those in CET 4 to determine their consistency. A total of 12 texts, 6 from Textbook 1 and 6 from CET 4's comprehensive reading section, are analyzed using Eng-Editor. The analysis is structured around three key dimensions: lexical, syntactic, and complexity grade. Findings reveal that Textbook 1's passages are more challenging than those in CET 4. Such insights can guide language educators in harnessing textbooks effectively to enhance students' language proficiency.
KEYWORDS: *Text Analysis; Complexity Evaluation; Eng-Editor; Textbook Comparison; Lexical and Syntactic Analysis.*

INTRODUCTION
In China, College English Test 4 (CET 4) and College English Test 6 (CET 6) are the high-stake exams to examine the English level and different abilities of university students. Reading comprehension is an important part in English tests, accounting for 35%, so that the score of reading section may affect the result of test. Textbooks can provide freshmen reading materials, which are much more difficult than texts in middle-school textbooks with more difficult vocabulary and complex sentences. Language teachers are willing to choose additional reading materials based on their experience to help learners acquire more vocabulary and improve reading abilities. Whether texts in textbook are suitable for freshmen? Whether can language teachers help freshmen improve language abilities and pass CET 4 with textbooks? How can language teachers choose other materials as additional materials?

Text complexity is used to judge the difficulty of text. This research collects 6 texts from Textbook 1 and 6 texts from the CET 4, and analyzes the text complexity from vocabulary, syntax, and complexity grade. The research will provide some suggestion for language teaching and learning to be aligned to China's Standards of English Language Ability (CSE).

1. LITERATURE REVIEW
1.1 Text complexity
Text complexity is multidimensional feature that consists of absolute and relative components included in language systems–phonological, morphological, syntactic, lexical (Bult´e & Housen, 2012). Text complexity analysis is the basis of test validation, reading test design, reading teaching and other relative studies. In 1970s, the studies focus on the text features, such as average length of sentence, average length of words and other features on surface. (Guo, Jin &Lu, 2018) The lexical complexity and syntax complexity may have great influence on the text complexity. (Alderson,2000)

1.1.1 Length
The length of text is an obvious index to test the complexity, but there are some other indexes, including vocabulary density, repetition of the words, conjunctions and so on. In CET 4, a reading text may consist of 300-350 words totally. The reading text consisting of 200-500 words is proper to be used in the test. (Liu & Han, 2000)

1.1.2 Vocabulary complexity
Vocabulary complexity is an important index of text complexity, and the dimensions include density, diversity, and sophistication. Lexical density is the ratio of the number of lexical (open-class) words to the total number of words in a text (Ure, 1971). The words can be divided into different classes, such as verb, noun, pronoun, adjective, adverbial, and other function words. If there are more notional words, the text is more difficult. The diversity of words is the range of vocabulary, so more formal words make the text more difficult. In China, China's Standards of English Language Ability (CSE) has been accepted as the criteria for judging students' language ability. College freshmen and sophomores should achieve Level 5, and they should master 5000—6000 words. Therefore, the higher vocabulary complexity is, the more words language learners should acquire.

1.1.3 Syntax complexity
The syntax complexity is another feature of text complexity. The syntax complexity of text reflects how productive various grammatical structures are (Foster & Skehan, 1996). The more accurate definition of syntax complexity is given in (Lu & Ai, 2015). They determine it as a part of linguistic complexity that studies the number and variety of syntactic structures and the degree of sophistication of those structures. The longer the sentence is, the more difficult it is. Texts with more sophisticated noun phrases and clauses must be more difficult to understand. Furthermore, text complexity can be reduced by adding more conjunctive words to connect

the sentences together, or close link among sentences.

1.2 Tool of testing

In the 1920s, the researchers began to conduct the studies on the text complexity. Flesch Reading Ease and Flesch-kincaid Grade Level (Kincaid et al. 1975) are always used to test the text complexity. In the 1990s, as the modern technology developed so quickly, researchers began to use computer and some software to analyze the texts and test the complexity. In China, Lu Xiaofei (2010) designed the L2 Syntactic Complexity Analyzer. Jin, T., Lu, X., Guo, K., Li, B., Liu, F., Deng, Y., Wu, J., & Chen, G. (2021) designed an online English text evaluation and adaptation system, Eng-Editor, which is popular in China. The online English text evaluation and adaptation system analyzes the text complexity from the following three dimensions: lexical analysis, syntactic analysis, and grade testing.

2. RESEARCH DESIGN
2.1 Research questions

College language learners have learned English for more than 10 years, but reading is the difficult part for them besides listening. It is common that language learners know meaning of every word in the passage, but it is difficult for them to understand it, especially doing some reading exercises of CET4/6. Especially for freshmen, texts of college English

textbook are difficult for large number of new words. For language teachers, textbooks are teaching materials, and it is also necessary to choose some additional materials to improve teaching and learning. However, most language teachers choose materials based on their own experience without any scientific standard. The research is to analyze text complexity of texts in Textbook 1 and texts in CET 4 with Eng-Editor. It is an easy way for language teachers to judge whether reading materials are suitable for language learners.

2.2 Research design

The research chooses 6 reading texts from College Textbook 1 and 6 texts from comprehensive reading section in CET 4 from 2019 to 2020 and analyzes the text complexity with Eng-Editor to examine consistency of text complexity between texts in Textbook 1and CET 4, including length of texts, vocabulary complexity, syntax complexity, the longest sentence, and complexity levels of texts.

3. RESEARCH RESULT
3.1 Research Analysis
3.1.1 Analysis of one text

Analysis of text complexity is conducted from vocabulary, sentences, and grade with Eng-Editor. The system can provide with more information. Take one text of CET 4 as an example. （Table 1）

Table 1. Lexical Analysis
(*Data coming from languagedata.net/tester*)

The lexicon in system consists of secondary-school lexicon, high-school lexicon, CET4 and CET 6 lexicons, and postgraduate lexicon. Vocabulary of the text is tested to belong to different stages and the percentage of the words occupying. Besides, the report may provide the frequency of occurrence in

the exams of different stages, repetition of each word, fraction of coverage and so on.

Table 2. Syntax Analysis
(*Data coming from languagedata.net/tester*)

Syntactic analysis is conducted from the number of qualitative subordinate clauses, complex noun clauses, non-finite verb phrases and attributive clauses. (Table.2) Qualitative subordinate clauses include definite clauses, gerund clauses, and noun clauses, etc. The complex noun phrases include noun phrases modified by adjectives, all-comers, prepositional phrases, determiners, present or past participles, etc.; "non-finite verb phrases include verb infinitives, gerund phrases, past participle phrases, etc. Attributive clauses include attributive clauses, adverbial clauses, and noun clauses. There are 9 qualitative subordinate clauses, 13 complex noun clauses and 7 non-finite verb phrases and attributive clauses in this 327-word text, and the longest sentence consists of 30 words.

Table 3. Text Grade
(*Data coming from languagedata.net/tester*)
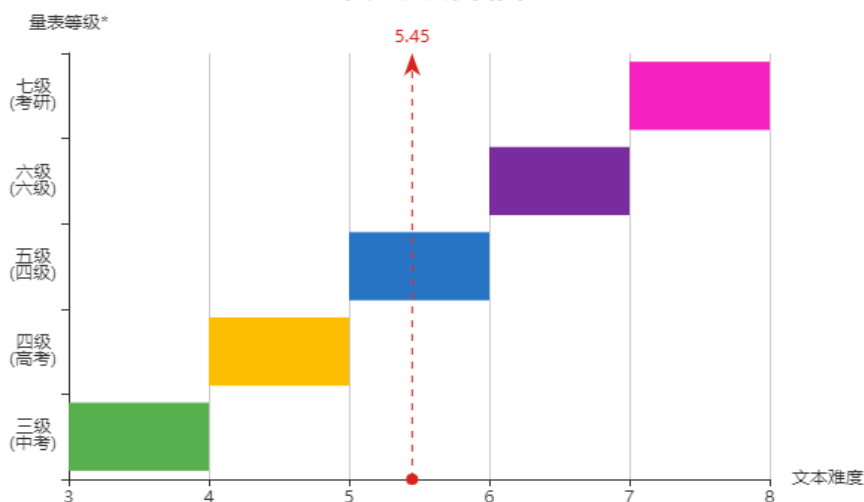
Based on the vocabulary and syntactic analysis, the system may provide a grade to match with the different levels of CSE. (Table 3) Vocabulary complexity of the text is 5.49, syntax complexity is 5.10, and text complexity is 5.45, at Level 5, which is consistent with CET 4 based on CSE.

3.1.2 Research Data
The research analyzes 6 texts in Textbook 1. The 6 texts are the first reading materials freshmen read in college English classes and language teachers must give instruction in detail. The complexity is consistent with texts in CET 4, which is good for language teachers to design reading tasks to help language learners improve language abilities.

Table 4. Text Complexity Analysis (Textbook)
(Data coming from languagedata.net/tester)

	T1	T2	T3	T4	T5	T6
Length(words)	614	675	818	794	702	750
VC/Grade	5.17/5	4.99/4	6.06/6	6.07/6	5.19/5	6.15/6
SC/Grade	3.98/3	5.47/5	5.17/5	5.95/5	7.30/7	6.23/6
Longest sentence(words)	42	44	34	56	44	43
LD/Grade	5.00/4	5.00/4	5.97/5	5.98/5	5.96/5	6.14/6
Level	5	5	5	5	5	6

6 Texts of Textbook 1 consist of 614-818 words, and are at Level 5, which is College English Level freshmen should achieve based on CSE. The vocabulary complexity (VC) is from 4.99 to 6.15. One text is in Grade 4, 2 texts are in Grade 5, and 3 texts are in Grade 6.

The syntax complexity (SC) is from 3.98 to 7.30, from Grade 3 to Grade 7. The longest sentence in each text consists of 34 to 56 words. The difficulty level (LD) is from 5.00 to 6.14, from Grade 4 to Grade 6. (Table 4) Although the longest text consists of 818 words, the longest sentence consists of 34 words, which can make language learners understand easily. Based on analysis of text complexity, texts in Textbook 1 are arranged according to difficulty and grade.

Table 5. Text Complexity Analysis （CET 4） *(Data coming from languagedata.net/tester)*

	T1	T2	T3	T4	T5	T6
Length(words)	327	349	321	335	360	330
VC/Grade	5.49/5	5.22/5	5.58/5	4.51/4	4.68/4	4.41/4
SC/Grade	5.10/5	5.29/5	5.42/5	5.05/5	4.73/4	3.89/3
Longest sentence(words)	30	30	35	35	46	47
LD/Grade	5.45/5	5.29/5	5.63/5	4.62/4	5.00/5	4.47/4
Level	5	5	5	4	5	4

In the CET 4 exams, all the reading texts consist of more than 300 words. Most high-frequency vocabulary have been acquired during the junior and senior middle school, but there are more words acquired during the college period. The vocabulary complexity (VC) is from 4.41 to 5.58, and the testing grade is on level 4 and level 5. The difficult of every text may on Lever 5, which can match with the grade of college learners mentioned in CSE. The syntax complexity (SC) is from 3.89 to 5.42. The longest sentence in each text also consists of more than 30--47 words. The difficulty level (LD) is from 4.62 to 5.63, at level 4 or level 5, and most texts are at level 5. (Table 5)

3.2 Research result
3.2.1 Analysis of texts in Textbook 1
The result indicates 5 texts in Textbook 1 are at Level 5, and the last text in this book is more difficult than previous ones at Level 6. Among these texts, 2 texts consisting of more than 600 words, 3 texts consisting of more than 700 words and one with more than 800 words.

The vocabulary complexity is from 4.99 to 6.15, from Level 4 to Level 6. The second text is the lowest and the sixth one is the highest. The texts are chosen from Unit 1 to Unit 6 with higher and higher vocabulary complexity. The syntax complexity is from 3.98 to 7.30, from Level 3 to Level 7. The first text is the lowest and the fifth one is the highest. The longest sentence is in the forth text with 56 words, and this text is the most difficult one among these 6 texts.

The longest text consists of 818 words, but the longest sentence consists of 34 words, which results in the complexity is below Level 6. The sixth text in Unit 6 is the most difficult one with the highest vocabulary complexity and the second highest syntax complexity. The texts are arranged mostly based on text complexity, which is consistent with the development of language learners' abilities.

3.2.2 Analysis of texts in CET 4
All the texts in CET 4 consist of more than 300 words, from 321 to 360 words. The vocabulary complexity is from 4.41 to 5.58, from Level 4 to Level 5. The syntax complexity is from 3.89 to 5.42, from Level 3 to Level 5. The longest sentence consists of 47 words, but it is not the difficult one for lowest syntax complexity.

The six texts are from 3 CET tests, and text 1 and text 2 are at Level 5 from the same test, which means comprehensive reading section in this test is more difficult than the same part in another two tests. To guarantee the validity of test, CET 4 at Level 5 is a standard test with a balance between vocabulary complexity and syntax complexity, which is consistent with CSE.

3.2.3 Comparison of text complexity

Table 6. Comparison of Text Complexity

	Length avg.	Vocabulary Complexity avg.	Syntax Complexity avg.	Longest Sentence avg.	LD/Grade avg.
Textbook	725.200	5.605	5.683	43.830	5.675
CET 4	337.000	4.982	4.913	37.160	5.077

The research compares text complexity of texts in Textbook 1 and texts of comprehensive reading part in CET 4. The average length of text in textbook is 725.200, and the average length of text in CET 4 is 337.000, which indicates there is a hug difference. However, all the texts are at Level 5. The vocabulary complexity and syntax complexity of texts in Textbook 1 is higher than texts in CET 4. The longest sentence of texts consists of more words than texts in CET 4.

Although texts in Textbook 1 are longer and more difficult, most of them are narrative essays, which makes language learners understand easily for coherence between paragraphs. All texts in CET 4 are expository and argumentative essays with greater logic.

3.3 Suggestions for language teaching and learning
Based on the data analysis of text complexity, the research can provide some suggestions for language teachers and learners.

In daily teaching, most language teachers choose the texts from some websites, exercises books and some test systems based on their experience. When they read the text, they judge the difficulty by the length of words, content, sentence structure and so on. If there are more new words they think or more complicated sentences in one text, they may pass it.

Textbooks, written by professors are important materials and texts are chosen according to analysis of text complexity. Language teachers should study textbooks from vocabulary, sentences, structure, and other characteristics of texts. Design of textbooks must follow some kind of principles; therefore, language teachers should have a deep understanding, then explore different teaching tasks. Selecting the original articles as additional materials can provide more reading materials for language learners, and language teachers should analyze text complexity of articles. If necessary, language teachers can lower text complexity by replacing difficult words with easier ones, rewriting sentences with less complex structure, or adding more conjunction words.

For language learners, textbooks are the basic materials with a large amount of vocabulary and complex sentences. Many learners do many reading exercises to pass CET 4 but ignore learning texts in textbooks. More exercises are good for improving proficiency in problem solving. Meanwhile, the coherence and cohesion of texts in textbooks can help them develop a logical thinking.

4. CONCLUSION
The research is conducted to compare the text complexity of texts in textbook and CET4. Texts in textbook 1 are more difficult than texts in CET 4, but all the texts are at Level 5, which is the standard level freshmen should achieve in CSE. The research also provides some suggestions to the language teachers and test designers.

The research provides the objective analysis on text complexity but lacks analysis from more dimensions, such as conjunctive feature of text, the average level of learners' language abilities, and so on. The text complexity can be an important factor to judge the difficulty of text. Language teachers and test designers should conduct more research on it and design the proper tests for the language learners to improve language abilities.

REFERENCES
1. *Alderson J C. Assessing Reading [M]. Cambridge: Cambridge University Press, 2000.*
2. *Bult´e, B. & Housen, A.. Defining and operationalizing L2 complexity. Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency In SLA. Amsterdam: John Benjamins Publishing Company. 2012, pp.21-46.*
3. *Foster, P., & Skehan, P. The influence of planning and task type on second language performance. Studies in Second Language Acquisition, 1996,18, 299–323.*
4. *Jin, T., Lu, X., Guo, K., Li, B., Liu, F., Deng, Y., Wu, J., & Chen, G. Eng-Editor: An online English text evaluation and adaptation system. Guangzhou: Language Data (languagedata.net/tester).2021.*
5. *Kincaid J P, Fishburne R p, Rogers R L & Chissom B S. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel [R]. Millington: Research Branch Repot of Naval Technical Training Command.1975*
6. *Lu, X., & Ai, H. Syntax complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. Journal of Second Language Writing,2015, 29, 16–27.*
7. *Ure, J. Lexical density: A computational technique and some findings. In M. Coultard (Ed.), Talking about text. Birmingham: English language research. University of Birmingham.1971, pp.27–48*
8. *郭凯，金檀，陆小飞.文本难度调控的研究与实践----从可读公式，多维特征到智能改编[J].外语测试与教学，2018，（3）：35-43.*
9. *刘润清，韩宝成.语言测试和它的方法[M].北京：外语教学与研究出版社，2000.*
10. *黄丽燕，王嘉樱.基于Python和Coh-Metrix 高考英语"一年两考"阅读文本复杂度分析【J】.外语测试与教学.2020,3,1-11.*
11. *中华人民共和国教育部.中国英语能力等级量表 [EB/OL].(2018-07-25). http://www.moe.gov.cn/s78/A19/yxs_left/moe_810/s230/201807/t20180725_343689.html.*