



Text Complexity of Reading Comprehension Passages in National Matriculation English Test: A Three-level Corpus Study

Yin Kailan

China

***Corresponding Author: Yin Kailan**, Exams, British Council, Chongqing, China. She actively engages in National Social Science Foundation Program and also is experienced in organizing large scale exams like IELTS and Aptis. Kailan.Yin@outlook.com

Abstract: This study investigated the text complexity of reading comprehension passages in China's National Matriculation English Test (NMET) of year 2020 and 2021, in the purpose of providing validation evidence for new NMET reform. Text complexity of 76 reading passages has been measured and compared on the three dimensions: lexical level, syntactic level, and discourse level. The natural language processing tools used in the study included Coh-Matrix and Eng-Editor. T-test and Wilcoxon test were conducted to compare the difference of each indicator.

The results suggested that the lexical level text complexity revealed the most evident changes between the two years. Significant elevation was found in lexical diversity of the NMET reading passages, in which the lexical diversity of 2021 NMET reading passages increased moderately compared with that of year 2020. The syntactic level text complexity also showed an inflation in noun phrases density in 2021 compared to that of 2020. Of the discourse level text complexity, insignificant increase of the indices occurred throughout the two years and the general trend was not necessarily rising. Nevertheless, the decrease of average hypernymy for verbs gave evidence of the growing text abstractness of NMET reading passages in 2021. Combined, the results might indicate that text complexity of the reading passages in the NMET from 2020 to 2021 has been steadily increasing by including low frequency and academic vocabulary, diversifying vocabulary in the passages, and complicating sentence structures. The results were further examined against the New English Curriculum Standards and guidelines to analyze whether the changes were reflected in the policies, and more often, of thematic context and genre, the passages of the two years' NMET employed unproportioned use of human and society and exposition. Suggestions for test designers and pedagogical practices were provided accordingly.

Keywords: National Matriculation English Test, text complexity, reading comprehension, Coh-Matrix, corpus-based study

1. INTRODUCTION

It has been 45 years since the resumption of Gaokao in 1977. Over the past years, great changes have taken place in all areas of China, with the economy advancing by leaps and bounds, culture becoming highly prosperous, and people's lives becoming increasingly affluent. This has laid a solid foundation for developing education in China, whereas also places higher demand on it. Within this frame of reference, the National College Entrance Examination, also known as Gaokao, a compelling examination that enrolls millions high school graduates, is being reformed constantly in accordance with the education requirements (Liu, 2017, 2019). In regards of English compulsory subject of Gaokao, the National Matriculation English Test (NMET) has also been taking advantage of language testing research, and considering the needs of enrollment system as well as the high school English teaching and learning, making timely adjustment to the content and form of NMET.

It is known to all, standardized examination was first introduced into China by the late Chinese famous linguist Professor Gui Shichun (1930-2017) who is also the first professional conducting successfully the ten-year (1990-1999) Matriculation English Test (MET) equating project sponsored by Ministry of Education of China (Gui, 1985, 2007, 2017). From 1978 to 1988, The National

Education Examinations Authority (NEEA) began to adopt one unified English test paper for Gaokao. Within a decade, the English subject of Gaokao varied its test structure every year, designed many word and sentence level questions and emphasized on examining the candidates' English Language knowledge rather than their ability to use it (Liu, 2017). In 1985, the Ministry of Education decided to conduct a standardized reform trial of Gaokao in Guangdong Province, initiating the standardized examinations in China. The forerunners Prof. Gui Shichun and Prof. Li Xiaojun contributed a lot to the standardization of English subject of Gaokao. For instance, they proposed five procedures to measure and maintain test standardization, which includes observing score distribution, item analysis, improving rating reliability, normalizing scaled scores and test quating (Li et al., 1989). The English subject of Gaokao was required to develop in a standardized way ever since, and was called Matriculation English Test (MET). In this period, the MET instruction was published by NEEA, multiple-choice questions began to dominate in MET, language skills and language use became the focus, and writing was introduced in MET (Qi, 2007). In 1991, the name National Matriculation English Test (NMET) was first employed and used to date. From 1991 to 1999, NMET included error correction, spelling and competing the dialogues, while reduced the number of grammar and vocabulary multiple choice questions. In 2000, NMET revolutionarily included listening in most provinces (Lv, 2017). In 2014, NEEA further adjusted the structure of NMET, replacing the previous single-sentence language use questions with discourse grammar fill-in-the-blank questions and launched a pilot in Zhejiang and Shanghai, whose NMET should include continuation writing and be conducted twice a year in 2016 (*A Year Two Test* reform). In 2021, Guangdong province, Jiangsu provinces and other 12 provinces/municipality adopted the *A Year Two Tests* mode.

During these reforms, reading comprehension still holds a dominating position in NMET, accounting for approximately 25% to 35% of the total NMET scores. The general NMET guideline made by NEEA requires the candidates to understand common topics, illustrate the main idea, structure and details, deduct the meaning of specific words and phrases and finally understand the opinions, purpose and attitude of the passages (Wang, 2018). However, few studies have investigated the textual characteristics of the passages after the *New NMET Reform*, whose new mode was employed by 14 provinces in 2021. Therefore, facilitated by the natural language processing tools, this study aims to examine the text complexity of the NMET reading comprehension passages from 2020 to 2021, in the purpose of providing validation evidence for New NMET Reform. It probes into the real condition in the text selection and presentation of the NMET reading test, and offers suggestions for the test developing and pedagogical activities.

2. LITERATURE REVIEW

2.1. Reading Comprehension and Text Complexity: Definition and Development

Successful comprehension of written reading assessment tasks is influenced by a variety of factors, such as the test taker's cognitive ability, knowledge, and motivation,; or test task characteristics such as task description, wording and format of questions, and context (Kintsch & Kintsch, 2005). In general, these factors can be grouped into three categories: reader, task, and text. Understanding the factors that influence reading comprehension can provide test researchers and educators with a deeper understanding of test development in a variety of academic areas, including reading, science, and social studies (Khalifa & Weir, 2009). In this study, the author puts a spotlight on the textual factors that influence reading comprehension that is text complexity.

Text complexity usually refers to the difficulty of a text and, in a narrow sense, equals to the linguistic features that affect text comprehension (Guo, et al., 2018). Research has shown that when text complexity is similar to the language level of foreign language learners, it helps to develop learners' language competency (Crossley et. al, 2012), while when text difficulty is much higher or lower than learners' language proficiency, it may hinder learners' language development (Kontovourki, 2012). Therefore, it is important to select text material of appropriate difficulty for learners.

Research on text complexity can be used to guide the design of reading tasks, daily classroom assessments, and to assess students' language proficiency in large-scale examinations (Lyashevskaya et.al, 2021). Previously, researches have measured text complexity with readability formulas such as the Flesch Readability the Flesch-Kincaid Grade Levels. Such formulas are easy to manipulate and

can visually detect text readability, but they examine the superficial text characteristics of the passages, and ignore syntactic, articulatory and semantic factors (Kintsch & Kintsch, 2005). The development of natural language processing and computational linguistics has prompted text complexity studies to incorporate deeper text features such as semantics, rhetoric, coherence, etc. (Guo, et al., 2018), typically represented by tools such as Coh-Matrix, Reading Maturity Matrix, and Text Evaluator. Jin et al. (2018) designed an Eng-Editor, a tool that could be used to evaluate and adopt texts based on the proficiency level specified in China's Standards of English Language Ability (CSE). This is the first tool that originated in Chinese English learning and teaching context, whose corpus is composed of The New English Curriculum Standards for Compulsory Education, The English Curriculum Standards for General High Schools (hereinafter called The New English Curriculum Standards), and the past NMET tests etc.

2.2. Researches on Text Complexity of NMET Reading Comprehension Passages

Part of the domestic studies on NMET reading comprehension focus on the reliability and validity of the content of reading comprehension questions in the college entrance examination co- temporally or over time (Gu, & Wang, 2008; Tao, 2017); the morph symbol ratio, lexical density, syntactic difficulty, and text length of NMET reading comprehension passages (Hu, 2018; Chen & Zhang, 2020). Most of these articles refer *The New English Curriculum Standards* and the syllabus for NMET. By Comparing the NMET reading comprehension with the two documents mentioned above from different perspectives, the researchers explore whether the development of NMET reading comprehension meets the requirements of the latter, or whether it is consistent with the documents, so as to judge the content validity of the tests (Xiao, 2014). However, the exploration of text complexity is not yet comprehensive. For example, studies mostly use the Flesch Readability of reading formula to calculate ease of reading, and the index has certain shortcomings. On the other hand, studies either start from vocabulary or syntax to explore text difficulty, and fewer combine the three aspects of vocabulary, syntax and discourse to explore comprehensively. Even fewer studies (Huang & Wang, 2020) have explored the text complexity of reading comprehension passages after the new NMET reform.

2.3. Lexical Level, Syntactic Level and Discourse Level Text Complexity

Word count and word length are the most direct measures of lexical level text complexity, which means that the more the number of longer words, the more difficult the text is to read, and studies have shown that it takes more time to process a longer word than a shorter word in English (Perfetti, 2011). It is also widely accepted in reading studies that readers with a large vocabulary could better understand the texts (RAND Reading Study Group, 2002). Inferring the meaning of a large number of new words based on a particular context will dampen learners' learning confidence (Far, 2016). Traditionally, lexical diversity has been calculated by type-token ratio (TTR). This formula with a larger type-token ratio indicating a more diverse vocabulary. Compared with TTR, Measure of Textual Lexical Diversity (MTLD) and VOCD, another two indices that could reflect the lexical diversity, were also less affected by the length of text (McCarthy & Jarvis, 2010). Lexical density is also an indicator of lexical complexity (RAND Reading Study Group, 2002).

Reading tests are often time-limited. Green et al. (2008) found that when time is limited, test takers become more stressed and their cognitive load increased accordingly. In general, longer sentences require more information to be processed and the accuracy of sentence comprehension is reduced (Far, 2016). Longer sentences and texts may also affect test takers' performance by reducing their working memory efficiency (Crossley et al., 2014). McNamara et al. (2014) argued that the shorter the sentence, the fewer words before the main verb or the fewer words before the noun phrase, the easier the syntax of the sentences is in the text; at the same time, readers may find the text more difficult when the density of passive voice sentences and negative sentences is too high. Latent semantic analysis (LSA) (Deerwester et al, 1990) and adjacent argument and stem overlap evaluation are effective way to decode the coherence of texts. LSA is "a mathematical method for computer modeling and simulation of the meaning of words and passages by analysis of representative corpora of natural text" (Landauer & Dumais, 2008). To construct a semantic space for a language, LSA first casts a large representative text corpus into a rectangular matrix of words by coherent passages, each cell containing a transform of the number of times that a given word appears in a given passage. The

matrix is then decomposed in such a way that every passage is represented as a vector whose value is the sum of vectors standing for its component words. Similarities between words and words, passages and words, and of passages to passages are then computed as dot products, cosines or other vector standing for its component words. Similarities between words and words, passages and words, and of passages to passages are then computed as dot products, cosines or other vector-algebraic metrics. Similarly, the Eng-Editor syntactic difficulty level will be encompassed to comprehensively examine the syntactic level text complexity of NMET reading passages.

Of discourse level text complexity, thematic contexts and genres, cohesion, text abstractness and readability are the four factors that influence the passages in this study. Specifically, all language learning activities should take place within a certain thematic context, for instance, and the New English Curriculum Standards provides 32 sub-themes based on different types of discourse and students should learn around such specific thematic context. Empirical studies have also shown that familiarity with a topic or genre influences test takers' performance on high-stakes tests (Crossley et al., 2012). While it is clear from the above discussion that word and syntax have an impact on reading, while the impact of the cohesion on reading remains controversial (Green et al., 2010). Cohesion refers to the specific elements of a text that indicates the coherent feature of the text and facilitates readers' comprehension (McNamara et al., 2014). A better understanding of the importance of cohesion in comprehension was the main inspiration for Crossley et al. to develop Coh-Metrix ("Coh" in Coh-Metrix means cohesion) (Crossley et al., 2014).

It is suggested that abstract text is more difficult to understand than content words or images in many researches (Corkill et al., 1988), possibly because when processing abstract text, readers' cognition is confined to a single language system (verbal or nonverbal), whereas when dealing with concrete language, readers can draw on knowledge of both linguistic and non-linguistic systems (concrete language may be pictorial) to aid comprehension (Green et. al, 2010).

The popular understanding of text complexity is approximately equivalent to readability or easeability. Readability formulas thus will be essential indicators in this study, and use Flesch Readability, Flesch-Kincaid Grade Level, and Coh-Metrix L2 Readability, together with Eng-Editor difficulty level to measure the discourse level text complexity of the NMET reading comprehension passages. The indices that will be used in this study are listed in Table 1 below.

Table1. *The adopted framework of text complexity*

Lexical level text complexity (L)
L1 Word count
L2 Average word length
L3 Average word frequency for content words
L4 Type-token ratio
L5 MTLTD
L6 VOCD
L7 Word beyond NMET syllabus
L8 Eng-Editor lexical difficulty level
Syntactic level text complexity (S)
S1 Average sentence length
S2 Noun phrase density
S3 Average modifiers per noun phrase
S4 Average words before main verb
S5 Agentless passive voice density
S6 Negation density
S7 Average argument overlap for adjacent sentences
S8 Average stem overlap for adjacent sentences
S9 Average LSA overlap for adjacent sentences
S10 Average LSA overlap for adjacent paragraphs
S11 Eng-Editor syntactic difficulty level
Discourse level text complexity (D)
D-TT Thematic context
D-G Genres

D-C1 Causal connectives incidence
D-C2 Logical connectives incidence
D-C3 Adversative and contrastive connectives incidence
D-C4 Temporal connectives incidence
D-A1 Average concreteness for content word
D-A2 Average hypernymy for nouns
D-A3 Average hypernymy for verbs
D-R1 Flesch Readability
D-R2 Flesch-Kincaid Grade Level
D-R3 Coh-Metrix L2 Readability
D-R4 Eng-Editor text difficulty level

3. METHOD

This section briefly introduces the research questions, research materials, instruments, and research procedure for analyzing the data.

3.1. Research Questions

- 1) What are the similarities and differences in the results of the lexical level text complexity of the NMET reading comprehension passages from 2020 to 2021?
- 2) What are the similarities and differences in the results of the syntactic level text complexity of the NMET reading comprehension passages from 2020 to 2021?
- 3) What are the similarities and differences in the results of the discourse level text complexity of the NMET reading comprehension passages from 2020 to 2021?

3.2. Research Materials

Considering that this study makes use of the method of text analysis, it is crucial to discern and clean the data used in this research.

The texts involved in this research are all texts extracted from the NMET reading tests from 2020 to 2021, which include multiple choices question type (four options given) and matching type (five out of seven items).

The dataset contains NMET reading comprehension passages from 2020 and 2021, which are 76 in number. The final word count of Year 2020 is 13474 and that of 2021 is 10685.

Table2. *Corpus of NMET reading comprehension passages*

Year	# of exams	#of passages	#word count
2020	9 (I, II, III, Q-I, Q-II, BJ, JS, TJ, ZJ)	43 (4+4+5+5+5+5+5+5+5)	13474
2021	7 (Q-I, Q-II, Q-Jia, Q-Y, BJ, TJ, ZJ)	33 (4+4+5+5+5+5+5)	10685
Total	16	76	24159

Note: *I, II, III, Q-I, Q-II, Q-Jia, Q-Y* is short for the national paper developed by NEEA. The difference is that, set I, II and III were used earlier than Q-I, Q-II, Q-J and Q-Y. Q-I, Q-II, were first used in 2020 and Q-J and Q-Y were first used in 2021. BJ, JS, TJ, ZJ is short for NMET reading comprehension passages developed by Beijing municipality, Jiangsu province, Tianjing municipality and Zhejiang province.

3.3. Instruments

The instruments used in processing the NMET test texts are Coh-Metrix¹ and Eng-Editor², which are two online text analysis tools and provide indices in lexis, syntax and discourse. R studio³ is an open source software for data analysis, as well as for producing charts and figures. Generally speaking, Coh-Metrix will provide the preliminary result of the 28 indices, such as word count and average word length. The other four indices: word beyond NMET syllabus, Eng-Editor lexical, syntactic and textual difficulty level will be extracted from Eng-Editor, and the last two indices, thematic contexts together with genres will be coded by the author manually according to the New English Curriculum Standards.

¹ cohmetrix.memphis.edu/cohmetrix2017

² <https://www.languagedata.net/tester/>

³ <https://www.rstudio.com>

3.4. Research Procedure

Specifically, after collecting the 76 passages of 16 NMET reading tests, the author processed the texts first, such as deleted title and subtitle, as well as the Chinese characters, and encoded them into Word. Then the texts were put into Coh-Metrix and Eng-Editor to extract the indices at lexical level, syntactic level and discourse level (see Figure 1)

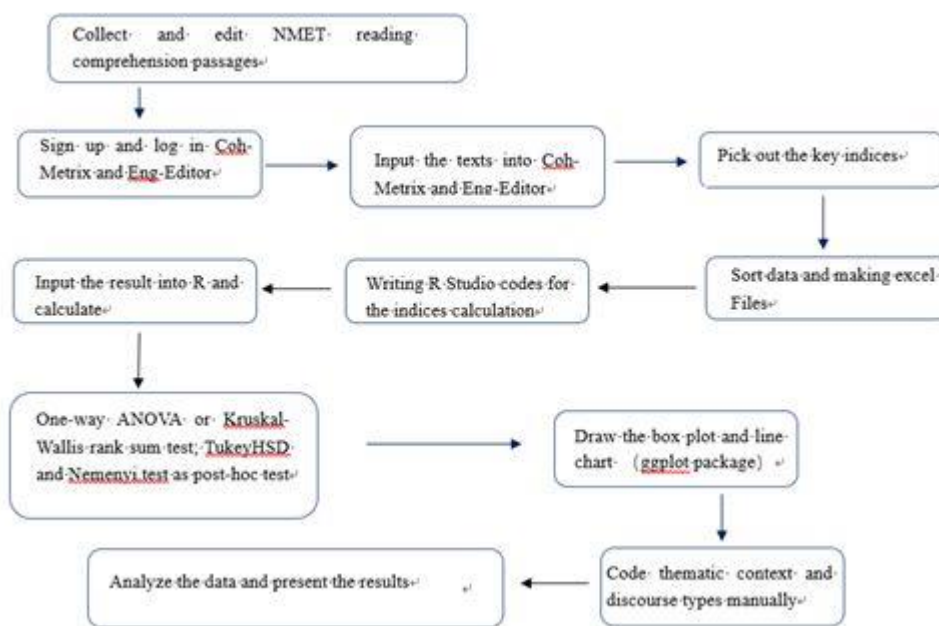


Figure1. Flowchart of the research procedure

The data were then put into excel files, because R Studio could not read file types such as word and txt. The values and groups were coded into “value” and “group”, so the code could run. Since there were two years' data to be compared, the test methods utilized in processing this group of data are T.test and Wilcox.test, depending on whether the data was distributed normally. In addition to the quantitative statistic, this thesis also included two qualitative measures, which were thematic contexts and genres. The author manually classified these two indices thrice, with a two-week gap each time. She also enquired her peers for help classify the two indices, so the correctness of this part is proved to some extent.

4. RESULTS

The three research questions will be addressed and the implication will be discussed in this section.

4.1. Research Question One: Lexical Level Text Complexity

The result of the eight lexical complexity indices is presented in Table 3. As shown, there is no statistically significant difference in the five indices: word count, average word length, average word frequency for content words, type-token ratio, and word beyond NMET syllabus. Meanwhile, statistically significant difference exists among lexical diversity, i.e. MTLN ($P=1.084e-13$), VOCD ($P=0.04$), and also Eng-Editor lexical difficulty level ($P=1.058e-13$), which will be discussed next.

Table3. The comparison result of lexical level text complexity indices of two years' NMET reading comprehension passages

	2020		2021		P
	Mean	SD	Mean	SD	
L1	306.68	78.35	314.88	85.51	0.99
L2	4.26	1.03	4.58	0.36	0.78
L3	2.16	0.30	2.26	0.17	0.37
L4	0.56	0.06	05.5	0.07	0.79
L5	98.83	25.18	101.36	21.37	1.084e-13
L6	100.04	22.85	104.45	23.37	0.04
L7	4.95	2.56	4.64	2.94	0.45
L8	4.23	0.50	4.64	0.49	1.058e-13

Table 3 and the Figure 2 show that the mean MTL D of reading comprehension passage of NMET 2020 is 98.83 and that of 2021 is 101.36. Therefore, it could be concluded that the reading passages of NMET 2021 have a higher MTL D ($P=0.04$). Thus, the lexical diversity of reading comprehension passages of NMET 2021 is higher than that of NMET 2020.

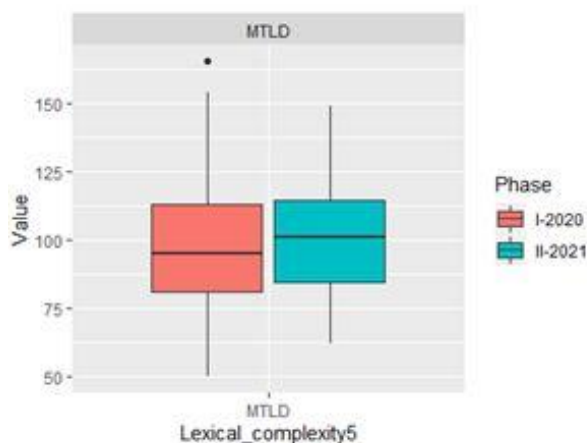


Figure2. Boxplot of MTL D of two years

Table 3 and the Figure 3 below present the result of VOCD of two years' NMET reading comprehension passages, which is another measure of lexical diversity. The mean VOCD of phase I is 100.04 and the mean VOCD of phase II is 104.45. Consistent with the result of MTL D, the VOCD of the two phases also appears to be statistically significant different ($P=0.04$). VOCD of reading comprehension passages of year 2021 NMET is also higher than that of 2020. Therefore, statistically, the lexical diversity of the NMET reading comprehension passages of 2021 is higher than that of 2020.

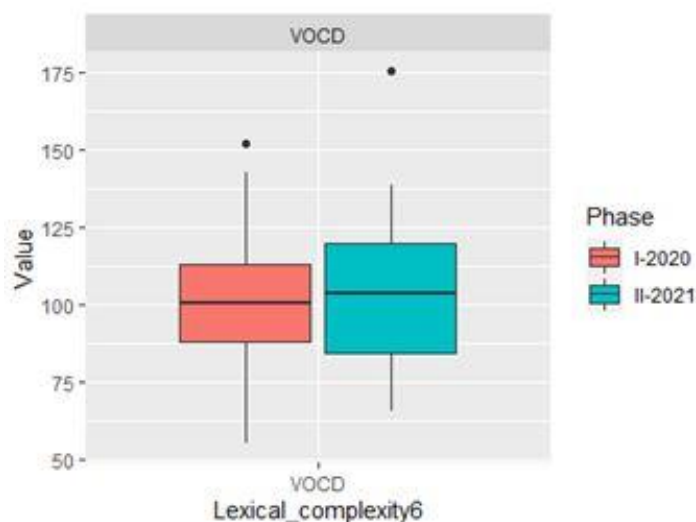


Figure3. Boxplot of VOCD of two years

In addition to MTL D and VOCD, statistically significant difference was also found in the index: Eng-Editor lexical difficulty level ($P=1.058e-13$). Overall, the mean value of Eng-Editor lexical difficulty level of year 2020's NMET reading comprehension passages is 4.23 and that of year 2021 is 4.64. The result may indicate that the Eng-Editor lexical difficulty level of 2021 is statistically higher than that of 2020. CSE has categorised nearly 3000 descriptors into 9 proficiency levels with 3 stages depicting the development of language ability (NEEA, 2018). Among which, candidates rated with level 1, 2 and 3 are at the elementary stage, while candidates of level 4, 5 and 6 are at intermediates stage and finally candidates at level 4, 5 and 6 belong to the advanced stage. So, according to the boxplot, for both two years, some NMET reading comprehension passages are over the level 6 or below 4, which

belongs to the difficulty level of the College English Test (CET) band 6⁴ and the difficulty level of National Senior High School Entrance Examination (NSHSEE) respectively. This result might denote that the difficulty level of the NMET reading comprehension passages have a relatively large range.

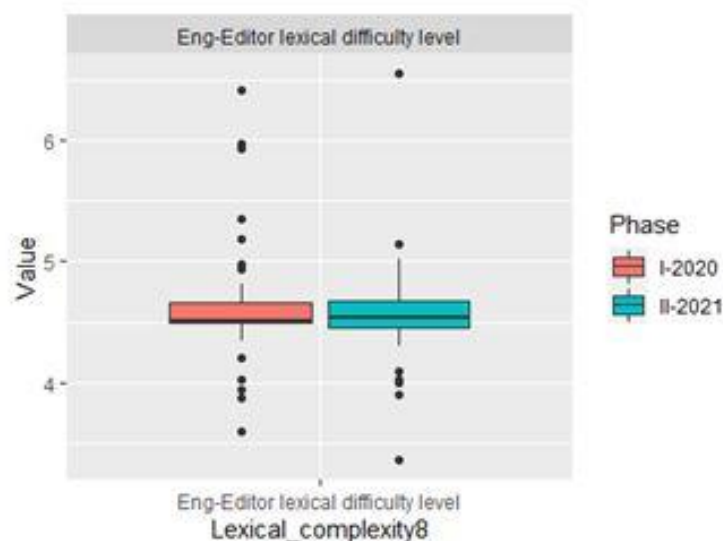


Figure4. Boxplot of Eng-Editor lexical difficulty level of two years

4.2. Research Question Two: Syntactic Level Text Complexity

After comparison, the result of the syntactic level text complexity of 2020 and 2021 is demonstrated in Table two. According to the data’s distribution type, the author found that there is only one indicator, noun phrase density performs statistically significant difference, while the other ten all show no statistically strong distinction.

Table4. The comparison result of syntactic level text complexity indices of two years’ NMET reading comprehension passages

	2020		2021		P
	Mean	SD	Mean	SD	
S1	16.06	3.52	15.50	3.47	0.64
S2	382.01	38.43	363.45	30.27	0.02
S3	1.29	0.18	0.88	0.17	0.46
S4	3.74	1.56	3.57	1.31	0.61
S5	5.82	4.82	6.19	6.12	0.84
S6	5.74	5.67	6.14	3.97	0.29
S7	0.47	0.15	0.41	0.17	0.13
S8	0.38	0.34	0.16	0.17	0.34
S9	0.17	0.07	0.16	0.06	0.25
S10	0.31	0.11	0.31	0.13	0.71
S11	4.05	0.67	4.32	0.60	0.008

Noun phrase density has shown statistically significant difference between year 2020 and 2021’s NMET reading comprehension passages. With the mean noun phrase density of 382.01 and 363.45 respectively, 2020’s noun phrase density is significantly higher than that of 2021 according to the significant test method (P=0.02). The boxplot demonstrates that the noun phrase density of NMET reading comprehension passages developed in 2020 is higher than that of 2021.

⁴ The College English Test, better known as CET, is a national English as a foreign language test in the People's Republic of China. It examines the English proficiency of undergraduate and postgraduate students in China. It includes two levels: CET4 and CET6 and enrolls millions of candidates each year.

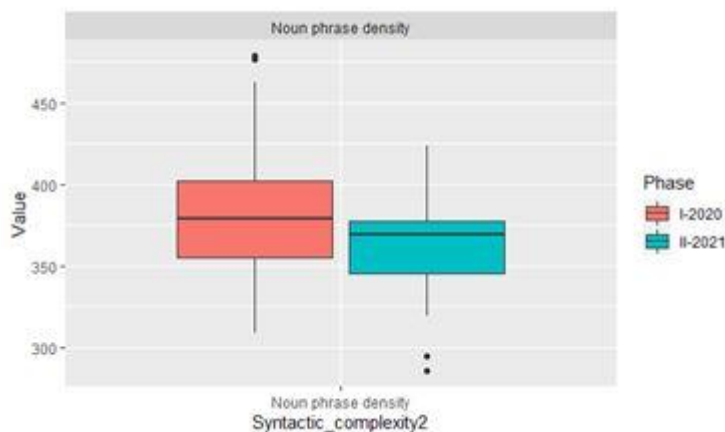


Figure5. Boxplot of noun phrase density of two years

4.3. Research Question Three: Discourse Level Text Complexity

Discourse level text complexity is composed of five parts: thematic context, genres, connectives, text abstractness and readability. After performing the significance testing method, the indicators show no statistically significant difference except average hypernymy for verbs ($P=1.081e-13$). In terms of thematic context and genres, because the dataset is not suitable for parameter test or non-parameter test, the author will present the descriptive analysis of these two sets of data.

Table5. The comparison result of discourse level text complexity indices of two years' NMET reading comprehension passages

	2020		2021		P
	Mean	SD	Mean	SD	
D-C1	23.62	9.96	24.54	9.16	0.68
D-C2	34.15	11.75	35.23	14.53	0.73
D-C3	13.93	8.35	15.51	9.06	0.45
D-C4	17.60	9.70	18.11	10.12	0.97
DA1	394.43	29.75	386.58	26.04	0.23
DA2	6.28	0.51	6.34	0.56	0.62
DA3	1.68	0.20	1.57	3.29	1.081e-13
DR1	63.47	9.86	63.10	12.72	0.92
DR2	8.71	1.96	8.25	2.35	0.70
DR3	15.51	5.99	14.70	4.35	0.50
DR4	4.59	0.41	4.57	0.44	0.48

Table6. The thematic context of NMET reading comprehension passages of two years

	Human and society		Human and themselves		Human and nature	
	2020	2021	2020	2021	2020	2021
BJ	2	2	2	2	1	1
JS	2	\	2	\	1	\
I	1	\	3	\	1	\
II	2	\	2	\	1	\
III	3	\	1	\	1	\
Q-I	3	2	2	2	0	1
Q-II	3	3	1	1	1	1
Q-J	\	2	\	1	\	2
Q-Y	\	4	\	0	\	1
TJ	2	2	2	2	0	0
ZJ	2	1	2	2	0	1
Sum	20	16	17	10	6	7
Total (76)	36(47%)		27(36%)		13(17%)	

Table 6 reports the thematic context involved in the reading texts of NMET 2017 to NMET 2021. Likewise, the proportion of the three categories of each phase is uneven. The overall trend is that in the two years that the proportion of human and society (47%) ranked first of thematic context. In contrast, human and themselves coupled with human and nature take account of 36% and 17% of the whole thematic context respectively. Nevertheless, the proportion of NMET 2007 in using the three categories of thematic context passages did not become better than that of 2020, with the proportion of human and society human and themselves still occupying most of the shares. All in all, the proportion of human and society is higher than the other two, especially human and nature, thus the distribution of the three categories being imbalanced.

Table7. *The genres of NMET reading comprehension passages of two years*

	Practical writing		Expository		Narration		Argumentation	
	2020	2021	2020	2021	2020	2021	2020	2021
BJ	1	1	1	2	2	1	1	1
JS	1	\	1	\	1	\	1	\
I	1	1	3	\	1	\	1	\
II	1	\	3	\	1	\	1	\
III	1	\	3	\	0	\	1	\
Q-I	1	1	3	2	1	2	0	0
Q-II	1	1	3	3	1	1	1	0
Q-J	\	1	\	3	\	1	\	0
Q-Y	\	1	\	3	\	0	\	1
TJ	0	1	0	1	0	1	0	1
ZJ	0	0	0	3	0	1	0	0
Sum	7	6	22	17	9	7	5	3
Total (76)	13(17%)		39(51%)		16(21%)		8(11%)	

The result in Table 7 indicates a common phenomenon of the reading comprehension passages the two years that the most frequently used genre is expository, taking up to 51%. The result might be attributed to the nature of NMET test, because it is such a large scale and high risk exam, and it has to be objective and avoid controversy, so expository is a “safe” choice. In contrast, argumentation tends to be least used in the five years’ NMET papers. While slight difference also exists among the genres involved in NMET reading comprehension passages of 2020 and 2021. For instance, the proportion of practical writing and narration is almost the same for two years (52%), but the proportion of argumentation of 2020 (11%) is higher than that of 2021 (9%).

Finally, the hypernymy for verbs is crucial in determining the abstractness of a passage. The hypernymy for verbs decreases from 1.68 to 1.57 (P = 1.081e-13) from Year 2020 to 2021, suggesting that the NMET reading comprehension passages of 2021 have become abstract verb concepts and therefore more difficult to understand from the perspective of text abstractness than the NMET reading comprehension passages of 2021. The boxplot below also shows the decreasing trend of this index from 2020 to 2021.

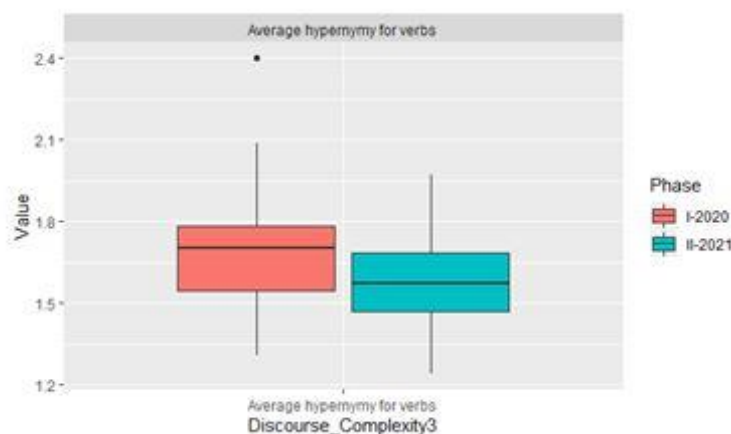


Figure6. *Boxplot of average hypernymy for verbs of two years*

5. DISCUSSION

In the preceding section, the author compared the text complexity of the NMET reading comprehension passages between Year 2020 and 2021. The results of comparing the lexical complexity of the reading texts of the two phases suggest that there is no statistically significant difference in their word count, average word length, average word frequency for content words, type-token ratio and word beyond NMET syllabus. To rephrase it, the most direct measures of lexical complexity, word count, word length and word frequency imply no strong distinction after the reform. Nevertheless, statistically significant difference has been found among MTLN, VOCD, and Eng-Editor lexical difficulty level. MTLN and VOCD are the major predictors of lexical diversity, and both the two indicators in phase II are higher than phase I; therefore, it is proper to say the lexical diversity of phase II increased after the NMET reform.

In terms of syntactic complexity, only the indicator, noun phrase density, evinces statistically significant difference in comparing the two years' data. The result might help to prove that the syntactic level text complexity of the two years' NMET reading comprehension passages is controlled reasonably. For example, the most direct indicator to measure syntactic level text complexity, average sentence length is found to fluctuate slightly over the past years, while overall the average sentence length of two years is 15 to 16. Only that the noun phrase density shows statistically significant difference in two years, the reason might ascribe to that in 2020, the NMET designers employed more passages with the topics of human and society, while using more expository to communicate this idea, while, these topics and genres naturally contains more noun phrases.

Thirdly, among the indices of discourse complexity, no statistically significant difference was found in the indices of connectives and readability. Nevertheless, the text abstractness index: average hypernymy for verbs shows statistically significant difference. Average hypernymy for verbs is an important indicator of text abstractness, and the decreasing trend of average hypernymy for verbs from 2020 to 2021 might hypothesize that the NMET reading comprehension passages in 2021 become more abstract. Combing the previous elaboration on noun phrase density, it could be summarized that there is a trend in the NMET designing process to choose and adopt the passages to make them more difficult to understand, and thus to assess the candidates' core literacy of English language.

Finally, the proportion of man and society of thematic context and expository of genres are still much higher than the other categories, the same as the proportion of the two in each year's reading texts. This imbalanced composition of these two categories should be improved in accordance of the *New English Curriculum Standards* and testing syllabus.

6. CONCLUSION

The major findings shed light on the significance of text complexity research of texts of NMET reading texts from multiple angles. The importance of reading comprehension in NMET would barely be weakened. Therefore, on the one side, the text complexity features, instruments and results spotted out in this study offer valuable empirical evidences for NMET designers and future study. One the other side, the high school teachers and students would also find the phenomenon this study found informative and therefore utilized in the daily learning and teaching.

Specifically, of NMET developing group, the selection and adaptation of texts are essential in the process of the whole test development. The NMET designers are recommended to take into account all three aspects of text complexity: lexical level, syntactic level and discourse level, and choose or adopt the NMET passages accordingly. For example, on the basis of this study, the author suggests when choosing the text for NMET reading it would be better to choose a text with a length of 225-300 words, a type-token ratio of 0.5, a MTLN and VOCD of 90-110 and a difficulty rating of Eng-Editor 4 to 5. Meanwhile, the syntactic complexity of the reading texts should increase the proportion of complex sentences and the low density of negative and passive voice sentences. The thematic context could have a more balanced frequency among the occurrences of human and themselves, human and society and human and nature, also with an appropriate increase in the proportion of argumentative essays. The NMET developing group could capitalize on the readability formula, for instance a text is more appropriate for NMET reading comprehension passages with the following parameters:

Flesh Readability at 50 to 60,

Flesh Kincaid Grade Level at 8 to 9,

Coh-Metrix L2 Readability at around 15 plus Eng-Editor text difficulty level at 4 to 5.

Meanwhile, the high school teachers of English are highly expected to carefully study the *New English Curriculum Standards*, the CSE and the NMET testing syllabus and other official documents, and make good use of the textbooks. At the same time, teachers could apply the traditional tools for calculating the readability of different passages, such as the Coh-Metrix and Eng-Editor, to determine text complexity of the texts used in teaching. Also, by that means, English teachers could choose appropriately graded reading texts for students, and cultivate students' reading proficiency and critical thinking. When teaching, the teachers could focus on examining the lexical diversity of the word the students mastered.

7. LIMITATIONS AND FUTURE STUDIES

There are two limitations due to realistic reasons. The data collected in thematic context and genres were coded manually. Although the author tried to conduct the coding process thrice, the data might be subjective to some extent. Secondly, the theoretical framework in this study is relatively inclusive; however, it has only measured a part of the features of text complexity, for instance, the factors of syntactic simplicity were not included in the framework. The main reasons contribute to that text features used in this study are too most closely related to text complexity and also because too many indices make it difficult to deal with all the texts.

Therefore, it is optimal for future studies to code the qualitative indices such as thematic context and genres to triangulate the results. In addition, future researchers could examine other indices to reveal a more integral picture of the text complexity of NMET, such as the syntactic simplicity and narrative it of passages. Conclusively, scholars are also encouraged to include diachronic datasets to enrich the research and verify the validity of NMET reform.

REFERENCES

- Chen, K., & Zhang, J. (2020). A study on ensuring the rating quality of the continuation task in the National Matriculation English Test. *China Examinations*, (12), 38-43.
- Corkill, A. J., Bruning, R. H., & Glover, J. A. (1988). Advance organizers: Concrete versus 236 abstract. *The Journal of Educational Research*, 82(2), 76-81
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115-135.
- Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? a computational and psycholinguistic investigation of text comprehension and text processing. *Reading in A Foreign Language*, 26(1), 92-113.
- Far, M. (2016). The effects of text type, text length and text difficulty on vocabulary retention through grossing. *The Journal of Language Teaching and Learning*, 6(1), 92-104
- Green, A., Ünaldi, A., & Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing*, 27(3), 1-21.
- Green, A., Ünaldi, A., & Weir, C. (2008). The cognitive processes of second language academic readers. LLAS Pedagogic Research Fund Project Report 11. Retrieved from: <http://www.llas.ac.uk/projects>.
- Gu, X., & Wang, Q. (2008). Content validity study of the NMET reading comprehension tests. *Examinations Research*, (3), 102-114.
- Gui, S. C. (1985). *Standardized testing - theory, principles and methods*. Guangdong Higher Education Press
- Gui, S. C. (2007). *Gui Shichun's anthology on English language education*. Foreign Language Teaching Research Press.
- Gui, S. C. (2017). *Selected academic papers of Gui Shichun*. Shanghai Foreign Language Education Press.
- Guo, K., Jin, T., & Lu. X. (2018). Research and practice in text adaption: From readability formulas and multidimensional analyses to data-driven adaptations. *Foreign Language Testing and Teaching*, (3), 35-43.
- Hu, X. (2018). A corpus-based quantitative analysis on linguistic features of reading comprehension section in College Entrance Examinations English Exam. *English Teachers*, (19), 40-42.

- Huang, L., & Wang, J. (2020). Investigating reading text complexity of NMET based on python and Coh-Metrix. *Foreign Language Testing and Teaching*, (3), 1-11.
- Jin, T., Lu, X., Guo, K., Li, B., Liu, F., Deng, Y., Wu, J., & Chen, G. (2018). Eng-Editor: An online English text evaluation and adaptation system. Guangzhou: LanguageData (languagedata.net/tester).
- Khalifa, H., & Weir, C. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.
- Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris., & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 73-92). IEA Publishers.
- Kontovourki, S. (2012). Reading leveled books in assessment-saturated classrooms: A close examination of unmarked processes. *Reading Research Quarterly*, 47(2), 153 - 171.
- Li, W., Gui, S.C., & Li, X. (1989). Statistical procedures lation English Test. *Foreign Language Teaching and Research*, (2), 27-39.
- Liu, Q. (2017). The development of the NMET over the past forty years. *China Examinations*, (2), 14-19.
- Liu, Q. (2019). The strategies for protecting test fairness in China: Taking Gaokao as an example. *China Examinations*, (11), 1-6.
- Lv, S. (2017). Reflections and theoretical analysis on the issues of forty years NMET in China. *Examinations Research*, 13(6), 59-65.
- Lyashevskaya, O., Panteleeva, I., & Vinogradova, O. (2021). Automated assessment of learner text complexity. *Assessing Writing*, 49, 100529.
- McCarthy, P.M., Jarvis, S. (2010). MTL-D, VOCD-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42 (2), 381-392.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. Q. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- National Education Examinations Authority. (2018). *China's standards of English language ability*. Ministry of Education of People's Republic of China.
- Perfetti, C. (2011). Decoding, vocabulary, and comprehension: The golden triangle of reading skill. In M. G. McKeown, L. Kucan, M. G. McKeown, L. Kucan (Eds.), *Bringing reading research to life* (pp. 291-303). Guilford Press.
- Qi, X. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education*, 14(1), 51-74.
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Education
- Tao, B. (2017). An exploration into item writing for the English achievement Test & NMET. *China Examinations*, (3), 25-44.
- Wang, Q. (2018). Interpreting the six major changes of the Senior High School English Curriculum 2017. *Foreign Language Education in China*, (1), 11-19.
- Xiao, Y. (2014). Research on content validity of reading comprehension in National Matriculation English Test Beijing (2010-2013). *English Teaching & Research Notes*, (1), 56-62.

Citation: Yin Kailan. "Text Complexity of Reading Comprehension Passages in National Matriculation English Test: A Three-level Corpus Study" *International Journal of Humanities Social Sciences and Education (IJHSSE)*, vol 9, no. s1, 2022, pp. 46-58. DOI: <https://doi.org/10.20431/2349-0381.09S1006>

Copyright: © 2022 Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.