

# Software Range Application in Test for English Majors Band 4 and Band 8

Yue Guo<sup>1</sup>, Xianwei Pu<sup>2\*</sup>

{121114011367@njust.edu.cn<sup>1</sup>, phil667@njust.edu.cn<sup>2\*</sup>}

School of Foreign Studies, Nanjing University of Science and Technology, Nanjing, China

**Abstract.** In this paper, using the corpus analysis software Range and data analysis software SPSS, the characteristics of vocabulary usage related to reading comprehension in the TEM4 and TEM8 from 2012 to 2023 (excluding 2020) are analyzed. By comparing the vocabulary usage in reading comprehension of the two exams, this paper tries to provide strategies for English majors to prepare for TEM4 and TEM8 reading comprehension in vocabulary. The research results are as follows: (1) there are more academic words in reading comprehension of TEM8 than in TEM4; (2) Pronouns, conjunctions and propositions are used more frequently than other word categories, such as modal verbs; (3) There are significant differences in the use of pronouns, conjunctions and modal verbs between TEM4 and TEM8.

**Keywords:** TEM4, TEM8, Software Range, Reading comprehension, Vocabulary

## 1 Introduction

TEM4 (Test for English Majors Band 4) and TEM8 (Test for English Majors Band 8) have gained growing recognition in China and worldwide in recent years as the sole national standard tests for English major students. TEM4 is designed for students at the foundational stage of English learning, while TEM8 is intended for those who have completed a four-year undergraduate program. These tests have been in use for many years, dating back to 1900, and they serve as the benchmark for evaluating English teaching and learning in China.

At the same time, reading comprehension is always used as the main means of acquiring a language. In China, second language reading ability of English is in great demand when the language of science and technology and advanced research as well as a global language [1]. The focus of the current paper is reading comprehension of the TEM4 and TEM8.

In this study, firstly, the characteristics of vocabulary use in TEM4 and TEM8 reading comprehension are analyzed by using Range. Then the two are compared using SPSS to provide some references on vocabulary for English majors to review and prepare for the examination, and to provide a reference for the teaching of vocabulary for English majors in universities.

### 1.1 Software Range

Range, a free corpus tool designed by Paul Nation, is available from the website [2]. Users can use Range to compare a certain text against different vocabulary lists so as to find out what words in the text are and are not in the lists, and to see what percentage of the words in the text

are covered by the lists. The software itself has three base lists, among which Word List One covers the most frequent 1000 words of English and Word List Two contains the second 1,000 most frequent words. As for Word List Three, it includes 570 words which don't belong to the first 2,000 most frequent words of English but are frequently encountered in senior high school and university texts from a variety of subjects. In total, there are 2,570 base words in the three-word lists. All the three base word lists include both the base forms of words and their derived forms. For example, the headword ADMIRE has the following family members: ADMIRING, ADMIRES ADMIRED, ADMIRABLE and ADMIRATION.

Words in the first two base lists comes from A General Service List of English Words (henceforth GSL) by Michael West [3] based on a corpus containing 5 million words while words in base list three are from Academic Word List (henceforth AWL) summarized by Coxhead [4] based on a corpus of 3.5 million running words.

At present, there are few studies on reading comprehension in TEM4 and TEM8, and the reading comprehension of TEM4 or TEM8 is mainly analyzed separately. Most studies mainly analyze the validity of reading comprehension, as well as the coping strategies of reading comprehension [5, 6, 7]. This paper compares the two levels of exams of reading comprehension, which provides a reliable reference for students to prepare for the exam.

One generally agreed principle for language learning is that the most commonly used words and typical patterns should be learned first. Therefore, as the main source of language input, it is necessary for students to learn vocabulary that often appears in exams. In order to find out the characteristics of vocabulary usage in TEM4 and TEM8, this study answers the following two questions:

- (1) What are the general characteristics of vocabulary used in reading comprehension of TEM4 and TEM8?
- (2) Are there significant differences in the different parts of speech used in the reading comprehension of TEM4 and TEM8?

## 2 Methodology

### 2.1 Data Collection and Procedure

Data for the study consists of two corpora of the reading comprehension of TEM4 and TEM8 (Table 1). By searching for information on the Internet and relevant documents and textbooks in the library, the author obtained the reading comprehension of real exams of TEM4 and TEM8 from 2012 to 2022, sorted out and made corresponding documents, which were used as research corpus. Among them, the 2020 exam was cancelled due to the COVID-19 epidemic.

**Table 1.** Corpora word counts

Reading comprehension	Number of words	Number of articles
TEM4	18666	10
TEM8	27837	10

As only texts of pure format can be recognized by Range, the first step is to transit the materials into pure text format. Reading comprehension download format is PDF document, which needs

to be converted into the common TXT format of corpus. First, the author converts PDF to DOC(X) format, and then to TXT format. Researcher used the online Convertio tool to convert the PDF document into a WORD document, kept only the reading comprehension articles and deleted the reading comprehension questions and options. Then, Word text was converted into UTF-8 encoded txt text in batches by using AntFileConverter software [8]. After the conversion, it was proofread to form a full-text corpus.

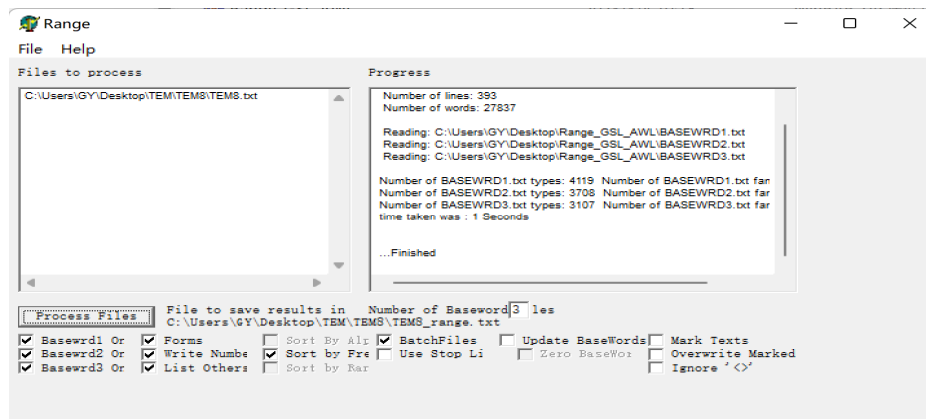


Fig. 1. The process of using Range

Using Range to analyze the text of TEM4 and TEM8. Range is designed based on word frequency, which is mainly used to analyze and compare the vocabulary of different texts [9]. After inputting the text of the research object into the Range, the tool of Sort by Fre was used to compares the recognizable words in the text with the three base word lists and calculates the number of words in the three base word lists and the frequency of the words appearing in the text in terms of the total vocabulary, as shown in Figure 1. In addition, the number of occurrences of each vocabulary and its ratio to the total vocabulary were also sorted in ascending order. Then, the data obtained from the Range analysis is imported into SPSS, and the function Analyze was used to conduct a comparative analysis of vocabulary of reading comprehension in TEM4 and TEM8, as shown in Figure 2.

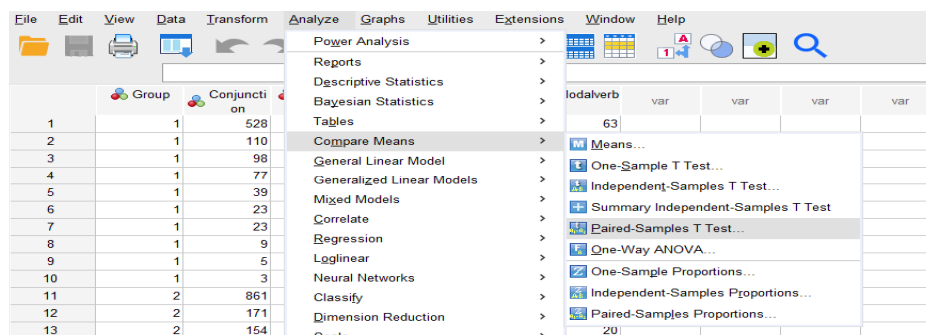


Fig. 2. The process of analyzing in SPSS

### 3 Results and Discussion

#### 3.1 Range Statistics of TEM4

The results showed that there are 4218 types in total, of which 1668 belong to the Basic Word List One, accounting for about 39.54% of the total vocabulary; 585 class symbols belong to the Basic Word List Two, accounting for about 13.87% of the total vocabulary; 363 class symbols belong to the Basic Word List Three, accounting for about 8.61% of the total vocabulary; Another 1602 words are outside the three basic word lists, accounting for about 37.98% of the total vocabulary (as shown in Table 2). These 4218 types appeared 18666 times in total.

**Table 2.** Range statistics of reading comprehension of the TEM4

Word list	Tokens/%	Types/%	Families
one	14348/76.87	1668/39.54	846
two	100/5.36	585/13.87	438
three	653/3.50	363/8.61	271
not in the lists	2665/14.28	1602/37.98	-----
Total	18666	4218	1555

The top five conjunctions are AND, AS, BUT, OR, IF. Conjunctions in discourse not only connect, follow and turn, but also introduce new information and topics, e.g. *And if someone offers you help, try accepting it instead of shunning it; But a growing body of research is suggesting the opposite is true* (TEM4, 2022). The pronouns that ranked high in word frequency were THAT, IT, I, YOU, HE, THEY, WE, SHE, and THEIR, e.g. *He can therefore take a more intelligent interest in his theoretical work* (TEM4, 2014). Their cumulative frequency was 6.9%, with the personal pronouns ranking high overall. The top prepositions based on frequency of occurrence are TO, OF, IN, FOR, WITH, ON, FROM, AT, e.g. *To attach the preparatory to the senior school these three classes were known officially* (TEM4, 2014). The data show that the notional verbs that appear more frequently are LIKE, SAID, MAKE, STUDY, READ, HELP. All of them are commonly used verbs to express a certain point of view or to introduce a new topic, e.g., LIKE, SAY. For example, *Skills like critical thinking and analysis must develop in the context of facts* (TEM4, 2015). Modal verbs with high frequency are CAN, WOULD, MAY, SHOULD, COULD and MIGHT. The frequency in reading comprehension of low and medium modal verbs in TEM4 is relatively high, while the frequency of high-value modal verbs (such as MUST) that strongly express personal opinions and feelings is relatively low. The total frequency of high frequency modal verbs is 186, but MUST only appeared eight times in the TEM4 corpus, e.g. *So literacy campaigns must be a good thing* (TEM4, 2017). This shows that the author's attitude in reading articles in TEM4 is more euphemistic and the expression is more objective.

#### 3.2 Range Statistics of TEM8

The results showed that there are 6143 types in total, of which 2022 belong to the Basic Word List One, accounting for about 32.92% of the total vocabulary; 905 class symbols belong to the Basic Word List Two, accounting for about 14.73% of the total vocabulary; 585 class symbols belong to the Basic Word List Three, accounting for about 9.52% of the total vocabulary;

Another 2631 words are outside the three basic word lists, accounting for about 42.83% of the total vocabulary (as shown in Table 3). These 6143 types appeared 27837 times in total.

**Table 3.** Range statistics of reading comprehension of the TEM8

Word list	Tokens/%	Types/%	Families
one	20863/74.95	2022/32.92	919
two	1675/6.02	905/14.73	602
three	1077/3.87	585/9.52	356
not in the lists	4222/15.17	2631/42.83	
Total	27837	6143	1877

The conjunctions with high frequency are AND, AS, BUT, OR, SO, IF. These conjunctions play the role of connection or transition, and construct the key vocabulary of reading texts, e.g. *And although I've been trying to reduce my meat consumption for health reasons* (TEM8, 2012). The data shows that the pronouns with the highest frequency are THAT, HE, IT, I, HIS, THEY and THEIR. Prepositions that appear frequently are OF, TO, IN, FOR, WITH, ON, AT and FROM, e. g. *It requires keeping track of changes around you in the same way that we monitor our surroundings when driving* (TEM8, 2014). Prepositions in reading comprehension articles usually appear in the form of prepositional phrases, while English prepositional phrases can play the role of adverbial, descriptive modifier and object complement in sentences, so prepositional phrases in texts can lead to ambiguity, so students should pay attention to prepositions when reviewing. The top notional verbs are LIKE, OWN, MAKE, COME, SAID, SEE, WENT, WORK, GO. The present tense and past tense of some verbs are among them, such as MAKE-MADE and SAY-SAID. The top modal words are can, WILL, WOULD, COULD, MAY and MUST. According to Halliday [10], CAN is a low-value modal verb, and its basic modal meanings include giving advice, giving permission, etc. WILL is a median modal verb with a strong modal meaning. MUST is a high-valued modal verb, and its modal meaning is very strong. The data show that the modal verbs of TEM8 are mainly middle and low value modal verbs, e. g. MUST only appears 13 times of total words, but CAN appears 48 times.

### 3.3 Comparison of TEM4 and TEM8

SPSS data shows that there is no significant difference in the type, token and word family of the reading comprehension in the three word lists, which shows that there is no significant difference in the distribution of the word in reading comprehension in the base word list. From Table 2 and Table 3, it can be concluded that the reading comprehension of TEM4 contains 8.61% academic vocabulary, and that of TEM8 contains 9.52% academic vocabulary. Therefore, students should pay more attention to academic vocabulary when preparing for the TEM8. And based on the data obtained from range, it is found that conjunctions, pronouns and prepositions are used more frequently than notional verbs and modal verbs in reading comprehension of TEM4 and TEM8, as shown in Table 4 (only the top ten words in each part of speech are selected for analysis).

**Table 4.** The usage frequency of different parts of speech in TEM4 and TEM8

	Conjunction	Pronoun	Preposition	notional verb	modal verb
TEM4	915	1181	1937	233	256
TEM8	1525	1722	2875	238	167

The usage frequency of conjunctions is high in both TEM4 and TEM8, with the top four being AND, AS, BUT, and OR. Using SPSS, the top ten conjunctions of TEM4 and TEM8 are tested by Wilcoxon signed-rank test (the data is not normally distributed). From Table 5, there is significant difference between usage of conjunction in TEM4 and TEM8 ( $Z = -2.8, p = .005$ ). In the use of conjunctions, there is a significant positive correlation between TEM4 and TEM8 (shown in Table 6). The reading comprehension passages in the TEM8 are relatively long, with a higher occurrence of long sentences guided by conjunctions.

**Table 5.** Wilcoxon Signed-rank test on the usage of conjunction in TEM4 and TEM8

	<i>N</i>	<i>M (SD)</i>		<i>Z</i>	<i>P</i>
		TEM4	TEM8		
Conjunction	10	91.5 (158.25)	152.5 (254.67)	-2.8	.005

**Table 6.** Pearson Correlations on the usage of conjunction between TEM4 and TEM8

		TEM4	TEM8
	Pearson Correlation		.999**
TEM4	Sig. (2-tailed)		<.001
	<i>N</i>	10	10

Paired sample T test was conducted on the pronouns in TEM4 and TEM8 (shown in Table 7). The data showed that there has significant difference in the use of pronouns between them ( $p < .001$ ). From Table 8, the Pearson correlation coefficient between them is significantly different ( $p < .001$ ). The reading comprehension of TEM8 is more difficult than that of TEM4, so pronouns are used more frequently. The pronouns of the top 10 in TEM4 appeared 1181 times, while those in TEM8 appeared 1722 times. The frequency of the demonstrative pronoun THAT ranks first in both TEM4 and TEM8. Therefore, students should cultivate sensitivity to personal pronouns and pay conscious attention to the backward references of pronouns when reading.

**Table 7.** Results of paired-samples t-test on the usage of pronoun in TEM4 and TEM8

	<i>N</i>	<i>M (SD)</i>		<i>MD</i>	95% CI	<i>t</i>	<i>p</i>
		TEM4	TEM8				
Pronoun	10	118.1 (58.05)	172.2 (80.84)	54.1	[-2.5, -.61]	-4.97	<.001

**Table 8.** Pearson Correlations on the usage of pronoun between TEM4 and TEM8

		TEM4	TEM8
	Pearson Correlation		.930**
TEM4	Sig. (2-tailed)		<.001
	<i>N</i>	10	10

Table 9 shows that there are differences in prepositions between the reading vocabulary of TEM4 and TEM8 ( $p = .019$ ). In the use of preposition, there is a significant positive correlation between TEM4 and TEM8 (shown in Table 10).

**Table 9.** Results of paired samples t-test on the usage of preposition in TEM4 and TEM8

	<i>N</i>	<i>M (SD)</i>		<i>MD</i>	95% CI	<i>t</i>	<i>p</i>
		TEM4	TEM8				
Preposition	10	193.7 (173.83)	197.7 (283.51)	104	[-1.63, -.14]	-2.85	.019

**Table 10.** Pearson Correlations Among on the usage of preposition between TEM4 and TEM8

		TEM4	TEM8
	Pearson Correlation		.987**
TEM4	Sig. (2-tailed)		<.001
	N	10	10

According to the data, there is no significant differences in notional verb between the reading vocabulary of TEM4 and TEM8 ( $p = .286$ ). In the use of notional verb, there is a significant positive correlation between TEM4 and TEM8 (shown in Table 11).

**Table 11.** Pearson Correlations on the usage of notional verb between TEM4 and TEM8

		TEM4	TEM8
	Pearson Correlation		.961**
TEM4	Sig. (2-tailed)		<.001
	N	10	10

Table 12 and Table 13 show that there are significant differences in the use of modal words between TEM4 and TEM8 ( $t = 4.9, p < 0.01$ ). The top five modal words in the reading section of TEM4 and TEM8 are probabilistic modal words, i.e., WILL, CAN, WOULD, MAY and SHOULD, accounting for 0.87% and 1.2% of the total vocabulary respectively. MUST appeared only 21 times in TEM4 and TEM8. This shows that compared with TEM4 reading, TEM8 reading uses more probabilistic modal words, and the author's attitude is more objective, so students may face more difficulties in judging the author's specific modal attitude.

**Table 12.** Results of paired samples t-test on the usage of modal verb in TEM4 and TEM8

	<i>N</i>	<i>M (SD)</i>		<i>MD</i>	95% CI	<i>t</i>	<i>p</i>
		TEM4	TEM8				
Modal Verb	10	25.6 (19.32)	16.7 (14.61)	8.9	[-.59, 2.47]	4.9	<.001

**Table 13.** Pearson Correlations on the usage of modal verb between TEM4 and TEM8

		TEM4	TEM8
	Pearson Correlation		.981**
TEM4	Sig. (2-tailed)		<.001
	N	10	10

## 4 Conclusion

In this study, the characteristics of vocabulary usage in TEM4 and TEM8 reading comprehension from 2012 to 2022 (excluding 2020) were analyzed and compared by using the corpus analysis software of Range and data analysis software SPSS. Based on the data obtained from range, it is found that conjunctions, pronouns and prepositions are used more frequently than notional verbs and modal verbs in reading comprehension of TEM4 and TEM8. All parts of speech in the TEM-4 and TEM-8 have strong correlation. And the vocabulary in the reading comprehension of TEM4 and TEM8 had small differences in prepositions, and significant differences in the use of pronouns, conjunctions, and modal verbs. It's not surprising to find differences in the use of pronouns, conjunctions, and modal verbs between these two levels, as they typically assess a wider range of language skills and proficiency. In general, TEM8 is expected to be more challenging and demanding in terms of English language proficiency compared to TEM4. According to the above findings, when English majors prepare for TEM4 and TEM8, they should consciously pay attention to the role of conjunctions in discourse, understand the various uses of prepositions, be mindful of the expression of modality and attitude in modal verbs, cultivate sensitivity to personal pronouns, consciously focus on pronoun reference and anaphora, and continually expand the breadth and depth of reading. In addition, although there is no significant difference in the use of notional verbs in the vocabulary of TEM4 and TEM8, the role of notional verbs in reading should not be underestimated, and special attention should be paid to the synonymy between notional verbs in the options and notional verbs in the passage.

## References

- [1] Ellis, R.: *Understanding Second Language Acquisition*. Oxford: Oxford University Press (1985)
- [2] Nation, P & Coxhead, A.: *RANGE* (2002)
- [3] West, M. A.: *general service list of English words, with semantic frequencies and a supplementary word-list for the writing of popular science and technology* (1953)
- [4] Coxhead.: *A new Academic Word List*. *TESOL Quarterly*, 34: 213-238 (2000)
- [5] Zhou, F. T.: *The Content Validity Analysis of the Reading Part of TEM-8*. 33:47-48 (2016)
- [6] Liu, L.: *The Content Validity Analysis of the Reading Part of TEM-4*. *Journal of Hubei University of Technology*. 28:124-128 (2013)
- [7] Zhong, Y. D.: *A study on coping strategies of reading comprehension in CET4*. *Foreign Language Teaching and Research*. 7: 113-116 (2018)
- [8] Anthony, L.: *AntFileConverter (Version 2.0.2) [Computer Software]*. Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software> (2022)
- [9] Bao, G & Wang, X.: *An Application of RANGE in the evaluation of productive vocabulary, Electronic Teaching of Foreign Languages*, pp. 4-58 (2005)
- [10] Halliday M. A. K & Hasan, R.: *Language, Context, and Text: Aspects of Language in a Social-semiotic perspective*, Victoria: Deakin University Press (1985)