A Diachronic Study on the Content Validity of the 2016-2019

——TEM-8 Section A Based on Corpus

Yu Zhang

School of Foreign Language, China West Normal University Nanchong, Sichuan province 637001

Absrtact: TEM-8 is an important high-risk examination for all senior English majors. By exploring the content validity of the section A part of the 2016-2019 TEM-8 reading, this paper aims to provide reference for the future direction of TEM-8 reading, and also to help students understand some of the contents of the TEM-8 reading examination in recent years to help them prepare for the exam. This paper mainly compares with the requirements of the TEM-8 test syllabus to explore the content validity of the reading article in the theme, genre, vocabulary density and length of the article in 2016-2019. It is concluded that in terms of theme and genre, the articles in recent years basically meet the requirements of the syllabus of the TEM-8 examination, but in the genre, the proportion of argumentative papers and narratives is on the high side, which should be deleted. Besides, this paper builds a corpus of reading articles in recent years, and uses LancsBox corpus retrieval software to make statistics on lexical type, token and type-token ratio. In terms of lexical density, the difficulty of articles in recent years basically meets the requirements of the syllabus of the TEM-8 examination, but the distribution of lexical density in 2017 is slightly unreasonable, so we should pay attention to avoid this phenomenon in the future. In terms of the length of the article, in addition to the uneven distribution of the text words in 2017, the length of the articles in this four years is reasonable, which meets the requirements of the total number of words in the syllabus of the TEM-8 examination and the number of words read per minute.

Key words: reading; theme; genre; corpus; lexical density; article length

Introduction

A corpus-based study has become a trend of research over the years. Alderson (1996) first brought corpus linguistics to the attention of language testing researchers[1]. And two basic scientific approaches of research: quantitative and qualitative analyses can be used in the corpus-based study. Quantitative analysis is frequency-based method. According to frequencies appearing in a certain corpus, linguistic features are classified, calculated, and analyzed statistically[2]. Besides considering that TEM-8 is a major test for all senior English majors, it is of great significance to the majority of English majors. Therefore, effective understanding of the content validity of TEM-8 can help students to prepare for exams better. Reading plays an important role in TEM-8, so this research aims to explore the content validity of the section A part of TEM-8 by building a small corpus and using the quantitative analysis. Because the syllabus of the TEM-8 examination was reformed in 2016, this research takes 2016 as the starting point to explore and 2019 as the end. I hope to provide reference for the future direction of TEM-8, and also help students understand some of the contents of the TEM-8 in recent years to help them prepare for the exam better. This paper is divided into five parts. In this paper, the theme, genre, lexical density and length of the reading articles from 2016 to 2019 are analyzed, and the results of content validity are obtained. Finally, the conclusion.

1. Theme

As can be seen from Table 1, the theme of the reading in 2016 includes three different aspects: literature, science and technology, education. And 2017 includes three aspects: life, education and society. Theme of 2018 includes two areas: society, scientific knowledge and literary reading, and 2019 reading test includes three themes: literature, biography and culture. From this, we can see that the theme of the 2016-2019 TEM-8 English reading examination covers a wide range of areas, including literature, science and technology, society, life, education, biography and so on, it meets the requirements of the syllabus of TEM-8, which requires the theme of reading should include society, science and technology, culture, economy, daily knowledge, biographies and some others. In addition, it can be seen from the reading articles in recent years that there are more articles on literature, and there is a slight deficiency in the theme of economy and daily knowledge. In general, from 2016 to 2019, the theme coverage of the TEM-8 reading articles is relatively extensive, and the theme of the articles in recent years all meet the requirements of the syllabus of the TEM-8 examination, thus the content validity is high.

Table 1: Statistics of reading theme and genre, 2016-2019

	PASSAGE	Topic	Theme	Genre
2016	PASSAGE ONE	The Great Gatsby	literature	narration
	PASSAGE TWO	Network security	science and technology	argumentation
	PASSAGE THREE	Higher education issues	education	argumentation
2017	PASSAGE ONE	Travel	life	narration
	PASSAGE TWO	Solutions of Extensive reading problems	education	argumentation
	PASSAGE THREE	National Service Projects	society	argumentation
2018	PASSAGE ONE	Immigrant	society	narration
	PASSAGE TWO	The Influence of Bilingualism on Personality	scientific knowledge	argumentation
	PASSAGE THREE	job-hunting	literature	narration
2019	PASSAGE ONE	Cover up your troubles	literature	narration
	PASSAGE TWO	Bennett and Gregory	biography	narration
	PASSAGE THREE	The bandit image	culture	expository writing

2. Genre

Also from table 1, we can see that genres of PASSAGE ONE, PASSAGE TWO, PASSAGE THREE in 2016 are narrative paper, argumentation and argumentation; And the three reading articles in 2017 are also narration, argumentation, argumentation; Articles of 2018 is narration, argumentation, and narration; The last three articles of 2019 are narration, narration and expository writing. It's not hard to see, All three articles from 2016 to 2019 focused on the proportion of narration and argumentation, The proportion of narration reached 50%, The proportion of argumentative papers has reached 42%, It occupies a large space in reading articles. In 2016 and 2017, the main articles were argumentation, and in 2018 and 2019, the main articles were narration, and the expository writing appeared in 2019 only, which indicates that the genre distribution of the TEM-8 reading articles in recent years was unreasonable. The reading section of the syllabus requires a variety of genres, including narrative text, descriptive text, expository writing, argumentation, advertisement, instructions, charts and so on. But from 2016 to 2019, the TEM-8 reading genres has failed to meet the requirements of various genres. However, since the TEM-8 examination is a test for senior English majors, and the argumentative paper in the senior student's academic examination occupies a heavy proportion, it

requires students to read and write argumentative papers effectively. Therefore, it is reasonable to occupy an important proportion of argumentation in the TEM-8 reading examination. In addition, the narrative texts related to literature, social life and other topics in line with the requirements of the syllabus for senior students, the syllabus requires students to be able to read difficult historical biographies and literary works, these narrations appeared in recent examinations are related to literature, social life and other topics, so it is also reasonable. However, the reading genre from 2016 to 2019 is relatively monotonous. Advertisements, instructions, charts and other genres may be involved in the examination so that the reading genres are more diversified while making the reading genres closer to the students' daily life, and helping students to better understand the genres of texts such as advertisements and instructions.

3. Word density

Text should be chosen as reading material with moderate difficulty and rich language information. Except for the facility value, TTR of text vocabulary plays an important role in the difficulty and richness of reading articles, TTR (type-token radio) is the ratio between words type and words token of the article. Token are all words form in a corpus, Type refers to the number of Token that are not repeated in a text. The type-token radio can be used to measure the lexical density of a text, because the type-token radio refers to the ratio between type and token in a particular corpus. As a result, the greater the TTR value, the higher the vocabulary variability, then the density of words increases. That is to say, the more difficult it is to understand the article [3](Chinese). This research takes lexical density as the criterion to judge the difficulty of the article, to build a small corpus of TEM-8 reading articles from 2016 to 2019, and the LancsBox corpus retrieval software is used to carry out the statistics of Type-Token ratio and the total Type-Token ratio (as shown in Table 2).

Table 2: Statistics of Type-Token ratio ,2016-2019

	PASSAGE	Token	Type	TTR	Total
2016	PASSAGE ONE	1171	552	47.14%	
	PASSAGE TWO	1148	583	50.78%	
	PASSAGE THREE	1123	488	43.46%	
Total		3442	1345	39.08%	
2017	PASSAGE ONE	887	439	49.49%	
	PASSAGE TWO	1422	512	36.01%	
	PASSAGE THREE	964	511	53.01%	
Total		3273	1193	36.45%	
2018	PASSAGE ONE	989	466	47.12%	
	PASSAGE TWO	1055	439	41.61%	
	PASSAGE THREE	1130	474	41.95%	
Total		3174	1096	34.53%	
2019	PASSAGE ONE	1117	550	49.24%	
	PASSAGE TWO	1159	517	44.61%	
	PASSAGE THREE	1060	485	45.75%	
Total		3336	1281	38.40%	
2016-2019	Total	13225	4915	TTR	37.16%

We can see from Table 2 that in the four years 2016-2019, except for the three articles (49.49%,36.01%,53.01%) in

2017, the TTR value gap of reading text in the other three years is not very large, and the text TTR values for 2016-2019 are 39.08%,36.45%,34.53%,38.40%, does not differ much from the total TTR value (37.16%) for the four years, which means the overall TTR value of the four-year reading text is balanced and the difficulty is balanced. Look at the Type-Token ratio every year, texts in 2016 were the highest in TTR during this four years, that is, questions that exist in 2016 reading texts are the most difficult (39.08%). And the TTR in 2018(34.53%) is the smallest, that is, the difficulty is the least. The difficulty of reading articles in 2017 is unreasonable, the second text is simple (TTR: 36.01%, less than 40%), and the third is obviously difficult (TTR: 53.01%, more than 50%). So this requires the author to consider the difficulty balance of each article in the future, try not to make the article too simple or too difficult, but to meet the requirements of the syllabus of TEM-8 examination. In view of the fact that this examination is aimed at the fourth grade students of English major, it is necessary to have some difficulty in the TEM-8 examination, and the difficulty of the reading text is balanced every year. In general, it meets the requirements of the syllabus of the TEM-8 examination, so the content validity is high.

Table 3: Statistics on the number of words read in texts, 2016-2019

	PASSAGE ONE	PASSAGE TWO	PASSAGE THREE	Total	words/min
2016	1198	1165	1154	3517	117
2017	911	1458	996	3365	112
2018	1002	1071	1131	3204	107
2019	1262	1245	1083	3590	120

4. Length of articles

Table 3 above shows the number of words read in every text and the total number of words per year for 2016-2019. As you can see, except for the uneven distribution of words in the 2017 text (911,1458,996). In 2016, 2018 and 2019, the number of words is around 1100, the length of PASSAGE ONE and PASSAGE THREE articles in 2017 is significantly shorter, under 1000 words, but PASSAGE TWO obviously has too many words, over 1400. Therefore, the presenter in the future should try to make the distribution of the article reasonable and uniform. Overall, more words in 2016 and 2019, above 3500. And the texts in 2017 and 2018 are close, all above 3200. The syllabus requires that the reading materials in TEM-8 should include around 3000 words. It can be concluded that the length of reading in 2017 and 2018 meets the requirements of the syllabus, while the reading length in 2016 and 2019 is obviously longer, exceeding about 500 words, which indicates that the total length of reading in 2016 and 2019 is unreasonable. This is also the next problem that TEM-8 reading should pay attention to. However, according to the statistics of the number of words that students can read per minute, the number of words read per minute in 2016-2019 is 117,112,107 and 120 respectively. All of which meet the requirement of the syllabus of TEM-8, which requires 50 words per minute. Considering that students still need some time to choose and fill in the answer sheet, so the reading speed of these four years is very reasonable, and its content validity is high.

5. Conclusion

In this research, through the form of tables, the author makes statistics on the theme, genre and so on, and draws relevant conclusions. In terms of theme and genre, the articles in recent years basically meet the requirements of the syllabus of the TEM-8 examination generally, but in the genre, the proportion of argumentative papers and narrations is on the high side, which should be deleted, and other article genres that students are required to master in the syllabus, such as advertisements, instructions, etc, should be increased. Besides, this paper builds a corpus of reading articles which appeared in recent TEM-8 tests, and uses LancsBox corpus retrieval software to make statistics on lexical Type, Token and Type-Token ratio. In terms of lexical density, the difficulty of articles in recent years basically meets

the requirements of the syllabus of the TEM-8 examination, but the distribution of lexical density in 2017 is slightly unreasonable, so the presenter should pay attention to avoid this phenomenon in the future. In terms of the length of the article, in addition to the uneven distribution of the three text words in 2017, the length of the texts in these four years is reasonable totally, which meets the requirements of the total number of words in the syllabus of the TEM-8 examination and the number of words read per minute. For the deficiency of reading texts emerged in recent years, I hope the presenter will pay attention to in the future and this article will provide some useful information for students to prepare for the examination.

References:

1. Sara T. Cushing. Corpus linguistics in language testing research[J]. Language Testing, 2017, 34(4): 441–449.
2. Liu Huibang. A corpus based study on the content validity of TEM8 passage correction [D]. Wuhan University of science and technology, 2010.
3. Wu Qi, Liu Lixiang. A corpus based study on the content validity of cloze in postgraduate entrance examination [J]. Examination research,2020,80(3):60-65.