

An Exploratory Study on the Easability¹ of Reading Tests for English Majors Band 4 (TEM-4) and Band 8 (TEM-8)

Shang-chun Zeng

College of International Exchange, Wuzhou University,
Wuzhou, Guangxi, CHINA
123209352@qq.com

Ying Shen*

Foreign Languages College, Guangxi University, Nanning,
Guangxi, CHINA
ying_kuomei@aliyun.com

Abstract—This paper aims to figure out the easability level of the reading texts in English Majors Band 4 (TEM-4) and Band 8 (TEM-8) from 2000 to 2014 by using corpus analysis software Coh-Metrix Version 3.0 and statistical analysis software SPSS 19.0. The easability features are analyzed with details in order to find some strategies and criteria for the selection of reading texts as well as the determination of text easability level in language testing. We hope that the results of this study can provide some theoretical database for the improvement of future language construct testing.

Keywords—*easability; Coh-metrix; TEM-4; TEM-8*

I. INTRODUCTION

Test for English Majors (TEM) is one of self-produced tests for testing English majors' English competence. It is divided into two levels: TEM-4 and TEM-8. The objective of TEM-4 aims to test English competence for second-year English majors while TEM-8 for fourth-year English majors in China. The overall purpose of these two tests is to check the language competence and performance of English majors at different stages, according to the requirements of Teaching Syllabus of Undergraduate English Majors of High Education Universities [7].

TEM-4 and TEM-8 have become the most important approaches to evaluating universities' teaching qualities and language proficiency of English majors in China. In the same time, reading abilities have become one of the important components of language teaching. Reading comprehension is a key part in both TEM-4 and TEM-8, and it takes twenty percent of the final scores. The requirements of Teaching Syllabus of Undergraduate English Majors of High Education Universities (2000) are that (1) students can read essays and comments in English newspapers as well as historical biography and literature works with difficulty published by English-speaking country; (2) students can catch the main ideas, text structure, language features and rhetoric uses in reading materials; (3) students reading rate should be 140 to 180 words per minute. Along with globalization and development of China's English teaching, there will be more requirements for English majors [6]. In order to reach above requirements for

understanding texts in TEM-4 and TEM-8, an analysis of the easability scores of reading texts of TEM-4 and TEM-8 may provide some implication for reading texts' selection and evaluation in terms of the future language construct testing .

II. RESEARCH METHODOLOGY

A. Research Subjects

This research takes section of reading comprehension in TEM-4 and TEM-8 from 2000 to 2014 as objects. There are six testing parts in TEM-4, they are dictation, listening comprehension, and language knowledge, cloze task, reading comprehension and writing before the reform in 2016. Five testing parts are included in TEM-8. They are listening comprehension, reading comprehension, language usage, translation and writing. Reading comprehension takes up a relatively heavy portion in both tests.

B. Research Methods

A small corpus was constructed by means of collecting data from Internet, library and related books. Two pieces of statistical analysis software (Coh-Metrix Version 3.0 and SPSS 19.0) were used to make a comparative study on the easability scores of reading texts in TEM-4 and TEM-8 from 2000 to 2014.

1) Coh-Metrix Version 3.0

Coh-Metrix is the abbreviation of Automated Cohesion Metric Tool based on computer technology. It mainly tries to make an automated evaluation of the cohesion of texts by means of over 200 indices of the linguistic and discourse representations of a text. Coh-Metrix can be used in many different ways to investigate the cohesion of an explicit text and the coherence of the mental representation of the text. [2]& [5].

Indeed, one motivation for the development of Coh-Metrix was to provide better measures of text difficulty [1]. Coh-Metrix gives eight principal easability components. These components provide a more complete picture of text ease (and difficulty) that emerge from the linguistic characteristics. The easability components provided by Coh-Metrix go beyond

¹ The meaning of "Easability" is Easier than Usable
More information from <http://jamin.org/easability-easier-than-usable/>

traditional readability measures by giving metrics of text characteristics on multiple levels of language and discourse. Moreover, they are well aligned with theories of text and discourse comprehension [3]&[4]. These easability components will be used to explain the exact text easability in TEM-4 and TEM-8 in this study.

The percentiles of each component as well as the z-score are used to evaluate the easability of each reading text in TEM-4 and TEM-8. A percentile score varies from 0 to 100%, with higher scores meaning the text is likely to be easier to read than other texts. For example, a percentile score of 80% means that 80% of texts are easier. A z-score is standard deviations a datum is above or below the mean, where the mean is set at 0. Generally, the z-scores are the preferred scores for research and statistical purposed, but the percentiles are more easily understood, particularly in a graph [5].

In this study, principal components analyses of TEM-4 reading tests are conducted first. After that the results of principal components in TEM-8 reading texts are presented. Besides, comparison between these two tests is made.

2) SPSS 19.0

SPSS, short for Statistical Package for Social Science, is one of the most widely available and powerful statistical analysis software packages. T-test in SPSS 19.0 will be used to evaluate the similarities and differences of the readability in TEM-4 and TEM-8 reading texts

C. Research Questions

The following research questions will be addressed.

- What are the characteristics of easability in reading comprehension texts of TEM-4 from 2000 to 2014?
- What are the features of easability in reading comprehension texts of TEM-8 from 2000 to 2014?
- Are there any differences in easability between TEM-4 and TEM-8 reading texts? And what are they, if there are?

III. PRINCIPAL COMPONENTS ANALYSES OF TEM-4 READING TEXTS EASABILITY

From 2000 through 2014, 60 texts of TEM-4 reading comprehension are analyzed, the mean length of each year four texts is 1,724 words ($SD=216.2$). The mean length of text is 431 words ($SD=54.1$).

TABLE I. PRINCIPAL COMPONENT ANALYSES OF TEM-4 READING TEXTS EASABILITY

TIME	PC1z	PC1p	PC2z	PC2p	PC3z	PC3p	PC4z	PC4p	PC5z	PC5p	PC6z	PC6p	PC7z	PC7p	PC8z	PC8p
2000	-0.63	26.76	-0.17	43.64	-0.09	46.41	-1.45	7.35	0.177	56.75	0.09	53.59	-3.34	0.04	0.41	65.91
2001	0.01	50	-0.19	42.86	0.5	68.79	-1.14	12.92	0.56	70.88	0.67	74.54	-3.26	0.06	0.55	70.88
2002	-0.17	43.25	-0.11	45.62	0.06	52.39	-1.32	9.34	1.83	96.64	0.23	58.71	-3.49	0.02	0.07	52.39
2003	-0.84	20.33	0.36	64.06	0.01	50	-1.47	7.21	-0.01	50	0.68	74.86	-2.80	0.26	-0.35	36.69
2004	0.11	53.98	-0.09	46.41	-0.67	25.14	-1.24	10.75	0.74	76.73	-0.09	46.81	-2.34	0.99	0.77	77.64
2005	0.17	56.75	0.29	61.03	-0.09	46.41	-1.24	10.75	0.42	66.28	-0.19	42.86	-2.29	1.1	-0.18	43.25
2006	0.43	66.28	0.46	67.72	-0.90	18.67	-1.28	10.03	0.14	55.57	0.30	61.41	-3.96	0	-0.46	32.64
2007	0.42	66.28	0.04	51.2	0.12	54.38	-0.85	20.05	0.80	78.81	0.72	76.42	-2.28	1.16	0.60	72.24
2008	0.24	59.10	0.38	64.8	-0.12	45.62	-1.65	4.95	-0.13	45.22	-0.19	42.47	-2.37	0.89	-0.65	26.11
2009	0.66	74.22	0.42	66.28	-0.03	49.2	-1.03	15.15	0.38	64.43	-0.53	30.15	-1.50	6.68	-0.14	44.83
2010	-0.49	31.21	0.16	56.36	-0.29	38.59	-1.24	10.75	0.49	68.44	0.37	64.06	-2.74	0.32	0.25	59.48
2011	0.13	55.17	0.22	58.71	0.01	50	-1.56	6.06	0.39	64.8	0.07	52.39	-2.04	2.07	-0.08	47.21
2012	0.14	55.57	-0.56	28.77	-0.12	45.22	-1.46	7.35	0.68	74.86	-0.13	44.83	-2.21	1.39	-0.51	30.85
2013	0.41	65.54	0.51	69.15	-0.64	26.11	-1.62	5.37	0.54	70.19	-0.43	33.36	-1.76	3.92	0.04	51.2
2014	0.02	50.8	-0.09	46.41	0.11	53.98	-1.27	10.2	0.55	70.54	-0.07	47.61	-2.35	0.94	0.23	59.1
MEAN	0.04	51.68	0.11	54.20	-0.14	44.73	-1.32	9.88	0.50	67.34	0.01	53.61	-2.58	1.32	0.04	51.36

Note: PC=principal component; z=z-score; p=percentile; PC1=narrativity; PC2=referential cohesion; PC3= syntactic complexity; PC4=word concreteness; PC5=deep cohesion; PC6=verb cohesion; PC7=connectivity; PC8=temporality

TABLE II. PRINCIPAL COMPONENTS ANALYSES OF TEM-8 READING TEXTS EASABILITY

TIME	PC1z	PC1p	PC2z	PC2p	PC3z	PC3p	PC4z	PC4p	PC5z	PC5p	PC6z	PC6p	PC7z	PC7p	PC8z	PC8p
2000	.35	61.45	-.40	34.65	.10	53.48	-1.38	10.62	-.14	44.65	.12	53.91	-2.15	5.43	.61	69.21
2001	-.16	44.05	-.56	30.28	-.52	34.10	-1.38	15.23	1.09	85.33	.79	75.13	-2.53	1.90	.32	62.38
2002	-.48	32.67	-.14	44.72	-.75	24.99	-1.52	7.77	.59	65.58	.34	62.24	-2.26	1.72	.78	76.66
2003	-.37	36.19	-.07	47.22	.98	83.02	-1.35	12.37	.12	54.56	.26	58.70	-2.09	3.59	-1.01	32.28
2004	-.03	48.98	.40	65.52	.57	70.15	-1.17	17.97	.65	73.35	.07	52.76	-2.90	.99	-.63	28.52
2005	-.23	41.79	.28	59.67	.53	67.74	-1.72	4.43	.19	56.27	-.25	45.17	-2.55	.86	-.52	30.64
2006	-.79	22.08	.02	50.60	.38	63.77	-2.10	1.88	.64	73.48	-.17	43.31	-3.61	.02	.29	59.86
2007	-.58	28.57	.01	50.40	.17	56.42	-2.00	3.56	.21	57.96	-.37	36.09	-1.82	3.53	.12	54.54
2008	-.48	32.21	-.13	44.84	.60	72.17	-1.71	5.18	.06	52.52	.32	61.94	-1.86	3.32	-.50	34.23
2009	-.23	40.99	-1.06	17.14	.36	63.18	-1.10	16.04	.33	62.22	.24	58.90	-1.13	12.99	-.64	40.47
2010	-.42	34.24	-.55	30.09	.29	61.21	-1.76	5.20	.45	65.59	.30	60.20	-2.30	2.08	.53	67.33
2011	-.35	37.84	-.05	48.03	.27	60.44	-1.62	5.27	-.45	32.96	.19	57.30	-2.58	.53	-.43	34.60
2012	.31	60.94	-.30	38.26	.46	67.65	-.68	25.00	.12	54.48	-.73	24.62	-2.08	4.96	-.13	44.93
2013	-.28	39.28	-.93	24.05	.45	67.02	-1.36	17.25	.45	67.17	.26	59.05	-1.91	3.85	.36	63.47
2014	.05	51.97	-.37	36.05	-.49	36.67	-1.58	6.13	-.23	41.40	-.38	35.62	-1.80	6.96	.19	56.22
MEAN	-0.25	40.88	-0.26	41.43	0.23	58.80	-1.50	10.26	0.27	59.17	0.07	52.33	-2.24	3.52	-0.04	50.36

IV. PRINCIPAL COMPONENTS ANALYSES OF TEM-8 READING TEXTS EASABILITY

From 2000 to 2014, a total number of 60 reading texts of TEM-8 are under analyses. The mean length of each year reading texts is 2,648 words (SD=316.90). The mean length of a text is 662 words (SD=79.23). The mean percentiles from PC1 to PC8 are 40.88%, 41.43%, 58.80%, 10.26%, 59.17%, 52.33%, 3.52% and 50.36% respectively. Connectivity seems to be the most difficult component, and syntactic complexity is of the least difficulty of eight principal components. Z-score of each component varies from -0.04 to -2.24 from 2000 to 2014. Among eight components, z-scores of PC7 and PC6 are more stable compared with other principal components. The mean z-scores of PC7 and PC6 are 0.07 and 0.04 respectively (see Table 2).

It is found that the mean percentiles of PC3, PC4 and PC7 of TEM-8 reading texts are lower than that of TEM-4 reading texts. PC3 here refers to syntactic complexity. The higher the score is fewer words, simple or familiar syntactic structures sentences may have. In the English language, there are occasions that sentences with fewer words tend to more difficult than those with more words. PC4 refers to word concreteness. When the score is higher, there is a higher percentage of content words, which are concrete in meaning rather than abstract in meaning. As a result, word concreteness score in TEM-8 reading texts is 14.07% higher than that of TEM-4. Besides, it may also be slightly influenced by the genres, such as narration and argumentation. PC7 means connectivity principal component. Higher score of this component indicates that a larger number of logic relations are

conveyed in details. Generally, it is acknowledged that reading texts in TEM-4 are easier than that in TEM-8 reading texts. Principal components analysis in Coh-Metrix provides probability to find out what aspects in TEM-8 reading texts are more difficult than reading texts in TEM-4.

After imputing all the data into SPSS 19, independent sample t-test is conducted to find the similarity and distinction. T-value may be above or below 0. Sig (2-tailed) is the probability of significance of t-test. Comparison results of these eight principal components are based on the Levene's test for quality of variances. The results are as follows:

First, Sig. of PC1 is 0.385 (z-score) and 0.307 (percentile) respectively, both are above 0.05 level. And thus, the variances of z-score and percentile of narrativity are equal. Taking into consideration of this, the significance probabilities for PC1z and PC1p are 0.043 and 0.037, which are lower than 0.05, as such, probability values are considered significant at the 0.05 level. As a consequence, there is statistically significant difference in narrativity between TEM-4 and TEM-8 reading texts.

Second, as for the second component (PC2: referential cohesion), the significance probabilities are 0.09 and 0.09.

Third, as to the third component PC3 (syntactic complexity), the significance probabilities are 0.022 and 0.012 ($p<0.05$) based on the z-score and percentile value.

Fourth, with regard to the fourth component PC4 (word concreteness), the probabilities of significance are 0.121 and 0.853; as such there is no significant difference between TEM-4 and TEM-8 reading texts in terms of PC4.

Fifth, considering the fifth component (PC5: deep cohesion), probabilities of significance are 0.146 and 0.098, both are more than 0.05. As a result, deep cohesion in reading texts of TEM-4 and TEM-8 shares no significant difference.

Sixth, thinking of component six (PC6: verb cohesion), the probabilities of significance are 0.811 and 0.802, far more than 0.05. Therefore, it can be said that there is no statistically significant difference in TEM-4 and TEM-8 reading texts' easability.

Seventh, with regard to component seven (PC7: connectivity), significance probabilities for z-score and percentile are 0.143 and 0.031. This means that for connectivity z-score, there is no significant difference.

Eighth, In relation to the last component (PC8: temporality), probabilities of significance are 0.656 and 0.865. As a consequence, no significant difference is found in reading comprehension texts' easability between TEM-4 and TEM-8.

Last, TEM-4 and TEM-8 reading texts from 2000 to 2014 are at a relatively stable level. However, the percentile of connectivity is in a significant difference between TEM-4 and TEM-8 reading texts

V. CONCLUSIONS

First, as for the eight principal components of text easability, TEM-4 and TEM-8 reading texts share some similarities as well as differences. In TEM-4 reading texts, deep cohesion appears to be the easiest. However, syntactic complexity is the easiest principal component in TEM-8 reading texts. Connectivity is the most difficult principal component both in TEM-4 and in TEM-8 reading texts' easability scores. This indicates that we should pay more attention to the acquisition of connectivity no matter what grade of language ability we may be in the future. The percentile and z-score of eight principal components are in normal distribution in both tests. There are statistically

differences in three components: narrativity, syntactic complexity and connectivity. For other five components, there is no significant difference.

Second, for the purpose of increasing the reliability of the easability testing in the future, easability can be conducted based on various types of genre, such as narration, exposition and argumentation and so on. Besides, under such micro data, random sampling can also be used to test validity.

In addition, easability level of the texts should match the language proficiency level of test-takers. It is the same as to the textbook editing. Each text in the book should be at an appropriate level so as to benefit language learners the most.

ACKNOWLEDGMENTS

This research is financially supported by Innovative Research Foundation for National University Students in 2014 (NO1410593049).

REFERENCES

- [1] Duran, N., Bellissens, C., Taylor, R., & McNamara, D.S., Qualifying text difficulty with automated indices of cohesion and semantics. In D.S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 233-238), Austin, TX: Cognitive Science Society, 2007.
- [2] Graesser, A. C., & McNamara, D. S., Computational analyses of multilevel discourse comprehension, *Topics in Cognitive Science*, 2011 (3): pp. 371-398.
- [3] Kintsch, W., *Comprehension: A Paradigm for Cognition*, Cambridge: Cambridge University Press, 1998.
- [4] McNamara, D.S., & Magliano, I.P., Towards a comprehensive model of comprehension, in B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 51, pp. 297-384), New York: Elsevier Science, 2009.
- [5] McNamara, D.S., Graesser, A.C., McCarthy, P. M., & Zhiqiang Cai, *Automated Evaluation of Text and Discourse with Coh-Metrix*, Cambridge: Cambridge University Press, 2014.
- [6] Yang H.Z, *Language Testing Theory and Practices course*, 2001.
- [7] Zou S. Experts Talking about TEM-4 and TEM-8, *Foreign Language Testing and Teaching*, 2012 (4): pp. 1-13.