



# Research on Data-Driven Text Adaptation of ESP Course Based on “*ENG-Editor*” Tool

Sun Rong<sup>(✉)</sup>

School of Foreign Languages, Guangzhou Xinhua University, Guangzhou, China  
nancysun2012@163.com

**Abstract.** The promotion of big data promotes the development of applied linguistics, and the cultivation of reading ability in foreign language teaching has always been the focus of research. In the process of learning ESP Courses, learners need to be provided with texts of appropriate difficulty for extended reading, which often requires the selection and adaptation of the original content of the text, that is, the regulation of text difficulty. Based on the situation of foreign language teaching and research in China, the study illustrates that text adaptation research is shifting towards the “data driven” stage by using the tool of “*Eng-Editor*”. Using “Wisdom Learning” as the theme, the research conducts a questionnaire survey of students from four aspects, which are vocabulary, grammar, difficulty, exercises, personal attitudes. Questionnaire survey and statistics are conducted with SPSS at the later stage to investigate the significance of the study and provide operational suggestions for future research.

**Keywords:** text adaptation · data-driven · *Eng-Editor*

## 1 Introduction

In the teaching of English as a second language, teacher-produced tests play an increasingly important role [1]. Teachers’ self-designed examinations are often used for a variety of purposes, such as recording students’ ability development, grading students’ level, and evaluating course teaching methods [2]. If English teachers can formulate high-quality questions, take into account the validity and reliability of the test, and scientifically measure students’ English ability, it will be helpful to rationally formulate the following teaching plan, and effectively promote the development of students’ English ability. Therefore, how to ensure the quality of English teachers’ self-proposed questions has gradually become an important topic for researchers [3, 4].

Most English teachers often encounter difficulties in the process of proposition, especially in the assessment of English reading ability. Providing reading materials with appropriate language difficulty for students is one of the many challenges faced by English teachers [5]. In order to accurately measure students’ reading ability, English teachers need to use reading materials that match students’ ability level. If the language difficulty is too high or too low, it will affect the final results of teaching effectiveness. Although there are many technical tools available to assist English teachers to complete

the difficulty regulation of reading materials, there are few studies on the difficulty regulation of ESP course reading texts. How to make intelligent assessment of reading texts is the focus of this research.

## 2 Literature Review

With the rapid development of information technology and corpus linguistics, there are many text difficulty control tools available at present, such as the Flesch-Kincaid Grade Level [6] and COH-Matrix [7]. This study is mainly based on the Eng-Editor developed by domestic researchers [5, 8], using the technology tools to assist teachers to quickly determine the reading material of language, especially vocabulary difficulty, and make a targeted adjustment. This research mainly on the empirical study, when choosing reading materials, the tools of technology based on the influence of core characteristics of the language difficulty, to quantify the degree of difficulty of text reading material, and through Eng-Editor establishes a difficulty prediction model for the learning of typical ESP reading text corpus, so as to help teachers complete the difficulty judgment of reading materials.

ESP course teachers need to judge and screen the language difficulty of the reading text. In many cases, the original text needs to be adapted to adjust the language difficulty to meet the test requirements, that is, text adaptation [9, 10]. In fact, most teachers mainly adjust the difficulty of reading texts based on their own teaching experience and subjective intuition. However, ESP courses are of various types, and teachers often compile test questions based on empirical intuition due to lack of unity objective standards, teachers use different language difficulties to read text, which will seriously affect the quality of test and test validity [11].

When adapting reading materials, technical tools provide teachers with quantitative reference ranges of text difficulty characteristic indexes at different levels through data mining of typical text corpora, identify the objects for adaptation and give feedback on adaptation suggestions, so as to assist teachers to complete difficulty adjustment of reading materials. Using the online learning platform for testing, collecting accurate and scientific primary data, establishing the corresponding large database, and accumulating semester data will be conducive to promoting teachers to make full use of difficulty control technology tools, which will improve evaluation practice and enhance teaching effectiveness.

## 3 Design of the Research

### 3.1 Research Object

The study objects are sophomore marketing major students in a university. Two parallel classes were randomly selected, one control class (43 students), one experimental class (45 students), the CET pass rate is 69.76% and 68.89% respectively, there is no significant difference in the scores.

### 3.2 Research Tool: *Eng-Editor*

Jin, T., et al. (2021) [12] developed the “Eng-Editor”, which firstly points out the difficulty level and secondly points out the difficulties. In order to point out the difficulty level, the “Eng-Editor” uses the original text in the “English test question Corpus” as the reference, conducts data mining on the original text, and establishes a difficulty level system matching the difficulty of the original text, including five levels of “high school entrance examination, college entrance examination, CET-4, CET-6, and Postgraduate Entrance examination”. When a new text enters “Eng-Editor”, it can be quickly matched to indicate the difficulty level. The word coverage rate is marked with different colors, and the number, proportion, density of superclass words, paraphrase of superclass words, and suggestion of replacement words can be pointed out.

In the process of intelligent adaptation, machine tools can provide data-driven support for foreign language teachers in three aspects, namely difficulty level judgment, quantitative adaptation suggestion and subsequent teaching feedback, to help complete difficulty adaptation of real texts and finally apply them to practice [8] (Lu & Jin 2017). On this basis, the steps of reading text difficulty regulation in this study are as shown in Fig. 1 and Table 1.

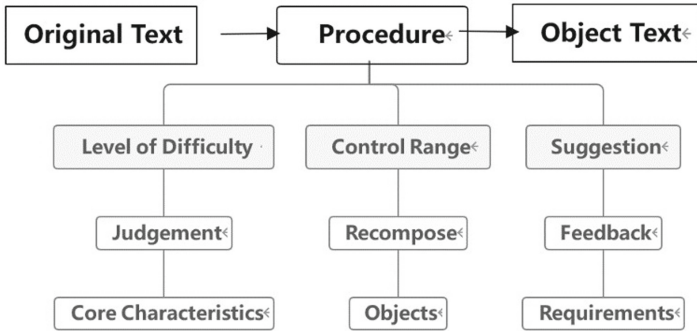


Fig. 1. Text Difficulty Adaptation Process

Table 1. Text Adaptation Index

Category	Words Analysis	Sentences Analysis	Discourse Analysis
Terms	Avg-syllables	Avg sentence length	Topic
	Avg-letters	Total clauses	Cohesion
	Total polysyllables	Total sentences	Coherence
	Total syllables		
	Total difficult words		
	Total words		

**Table 2.** Group Statistics

	Class	N	Mean	Std. Deviation	Std. Error Mean
Score	1 = Control Class	48	70.56	6.773	1.365
	2 = Experimental Class	45	78.09	5.832	1.194

In terms of vocabulary difficulty, the selection of articles with relevant materials according to the theme of the text is a crucial step to cultivate the reading ability, because neither too difficult nor too easy is conducive for students' study. In addition, the vocabulary difficulty of the text was evaluated by using "Eng-Editor", and the proportion of vocabulary of the CET-4 and CET-6 levels was controlled to be less than 3%. In terms of syntactic difficulty, the adaptation should take into account of word frequency and usage environment, especially whether ESP vocabulary is commonly used. It can focus on two aspects: firstly, the simplicity of syntax. Most of the texts in ESP teaching are expository or argumentative, and there are many professional terms. Secondly, Cohesion and coherence. Deep cohesion mainly reflects the degree to which the text uses causal and intentional conjunctions to express causal and other logical relations. These conjunctions can help learners understand events and behaviors. This study attempts to compensate readers in terms of connectedness.

### 3.3 Data Collection and Analysis

Based on the theme of intelligent reading, this study conducted a questionnaire survey on students to analyze the difficulties in reading, such as vocabulary difficulty, grammar difficulty, exercise difficulty, self-evaluation. Descriptive statistical methods were used to analyze the results to explore the difficulty of text difficulty adaptation and reading difficulty. At the same time, qualitative analysis method was used to analyze the result data of open questions, so as to verify or supplement the results and findings obtained by descriptive statistical analysis.

## 4 Results and Discussions

### 4.1 Analysis of Test Score

The result data of final examination was input into SPSS16.0 to sort out the statistics. Firstly, the author conducted ANOVA analysis based on reading scores and the results are shown in Table 2.

Table 2 shows that the average score of Experimental Class ( $M = 78.09$ ,  $Std. = 5.832$ ) is higher than that of Control Class ( $M = 70.56$ ,  $Std. = 6.773$ ), with a smaller standard deviation.

As shown in Table 3, the significance level of homogeneity of variance test was  $P = 0.063 > 0.05$ , indicating homogeneity of variance. In the study, there was a significant difference between the two learning effects ( $M = -6.741$ ,  $P = 0.000 < 0.01$ ). Experimental Class has a better learning effect and is conducive to improving students' English reading level.

**Table 3.** Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper	
Score	Equal variances assumed	3.310	.063	-3.867	87	.000	-6.741	1.898	-10.309	-3.759
	Equal variances not assumed			-3.867	82.738	.000	-6.741	1.898	-10.311	-3.756

\*\* p < 0.01

**Table 4.** Statistics of Self-assessment Score of Class A

	Abilities	M	SD
Vocabulary Difficulty	1. How difficult is vocabulary comprehension	3.05	0.98
	2. Whether the number of superclass words is reasonable	3.00	1.04
Grammar Difficulty	3. Whether the number of long and difficult sentences is reasonable	3.25	1.05
	4. Do long and difficult sentences affect understanding very much	3.15	1.08
	5. Whether the text length is acceptable	3.75	1.04
Exercises Difficulty	6. Whether the length of reading reasonable	3.35	0.98
	7. Whether the difficulty of the exercises is acceptable	3.25	0.97
Self-evaluation	8. Whether it is necessary to do extracurricular extended reading	4.05	0.91
	9. Is it helpful to enhance the understanding of the ESP topic	3.95	1.07
	10. Whether it helps improve reading ability	4.01	1.01

**4.2 Analysis of Self-assessment Questionnaire**

The self-assessment questionnaire involves four types of questions: “Vocabulary Difficulty”, “Grammar Difficulty”, “Exercise Difficulty” and “Self-evaluation”, with a total of 10 questions.

Table 4 shows that the average value of the two items of “Vocabulary Difficulty” is about 3, indicating that the overall score of students on vocabulary difficulty is moderate. The mean value of the three items of “Grammar difficulty” is above 3, which is slightly higher than the mean value of “Vocabulary Difficulty”, indicating that the grammar difficulty is slightly higher. However, for the two items of “Exercise Difficulty”, the standard deviation was also the smallest, indicating the low degree of dispersion of the score. The average value of the three items of “Self-evaluation” is about 4, which is the highest value among all the items, indicating that students have a positive attitude towards extracurricular reading and can effectively improve their reading ability.

## 5 Conclusion

According to the “Eng-Editor” tool, the difficulty of reading text in ESP course is regulated, which can effectively improve students’ reading ability. It can statistically mark vocabulary and generate vocabulary list. In the process of data-driven adaptation, teachers can neither fully expect students to find and consolidate knowledge by themselves, nor can they directly present vocabulary knowledge to students. In other words, after the teachers select and process the corpus, they present it to the students in the way of theme expansion, and the learning effect will be better. In the later research, relevant vocabulary teaching activities were designed based on the relevant functions of “Word Master App”, and activities such as “weekly word list” and “daily sentence making” were set up by ESP using big data. In the future, qualitative research methods or hybrid research methods can be adopted, and data collection tools such as talking thoughts and personal interviews can be used to further explore the difficulty regulation of texts and adapt them more scientifically, so as to help students more effectively and pertinently read ESP-related texts.

**Acknowledgement.** Funding: 2021 Shanghai Foreign Language Education Press and China Research Center for Foreign Language Teaching Materials and Teaching Methods: Research on the Application of Case Analysis Based on Ideological and Political Curriculum in ESP Teaching (NO. 2021GD0282).

## References

1. Anderson A, Schlueter J E, Carlson J F & Geisinger K F. Tests in print IX: An Index to Tests, Test Reviews, and the Literature on Specific Tests [M]. Lincoln, NE: University of Nebraska Press, 2016.
2. Arthaud T J, Vasa S F & Steckelberg A L. Reading assessment and instructional practices in special education [J]. *Diagnostique*, 2000, 25: 205–27.
3. Coniam D. Investigating the quality of teacher-produced tests for EFL students and the effects of training in test development principles and practices on improving test quality[J]. *System*, 2009(37), 226–42.
4. Graesser A C, McNamara D S & Kulikowich J M. Coh-metrix: Providing multi-level analyses of text characteristics[J]. *Educational Researcher*, 2011, 40:223–234.

5. Green A & Hawkey R. Re-fitting for a different purpose: A case study of item writer practices in adapting source texts for a test of academic reading [J]. *Language Testing*, 2012(29), 109–29.
6. Jin T, Guo K, Mak B, & Wu Q. Lexical profiles of reading texts in high-stakes tests: Where are the benchmarks? [J]. *IJCALLT*, 2017, 34–49.
7. Jin T, Lu X. A data-driven approach to text adaptation in teaching material preparation: Design, implementation and teacher professional development[J]. *TESOL Quarterly*, 2018, 2: 457–67.
8. Jin T, Lu X & i J. Syntactic complexity in adapted teaching materials: Differences among grade levels and implications for benchmarking[J]. *Modern Language Journal*, 2020(104), 192–08.
9. Jin T, Lu X, Guo K, Li B, Liu F, Deng Y, Wu J, & Chen G. (2021). Eng-Editor: An online English text evaluation and adaptation system. Guang-zhou: Language Data ([languagedata.net/tester](http://languagedata.net/tester)).
10. Jin T, Li Y, & Li B. (2016). Vocabulary coverage of reading tests: Gaps between teaching and testing[J]. *TESOL Quarterly*, 50(4), 955–964.
11. Jin Y. The place of language testing and assessment in the professional preparation of foreign language teachers in China [J]. *Language Testing*, 2010 (27), 555–84.
12. Kincaid J P, Fishburne L R P, Rogers R L & Chissom B S. Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel [R]. Research Branch Report 8-75, Millington, TN: Naval Technical Training, U.S. Naval Air Station, Memphis, 1975.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

