A Study on the Readability of Chinese Text from a Rhetorical Perspective

Dong Su[1,2], Zhou Jianshe[1,2(✉)], Zhang Kai[1,2], Zhang Wenyan[2,3], and Lu Yumei[3]

[1] Capital Normal University, Beijing 100048, China
2210101038@cnu.edu.cn
[2] Research Center for Language Intelligence of China, Beijing 100048, China
[3] China Fire and Rescue Institute, Beijing 102202, China

Abstract. Text readability calculation is an important method to evaluate the difficulty of Chinese reading text. This paper based on rhetoric corpus, counted the rhetoric features related index, adopt quantitative means to analyze the correlation of rhetoric features and text readability and other level language features fusion, and then use the multiple linear regression analysis method, with each level text features related index as an independent variable, the text difficulty level as dependent variable, to construct text readability calculation formula model, and calculated the model effect fitting degree. The results of this paper show that there is a strong correlation between the proportion of common rhetorical features, rhetorical questions, references and the readability of the text, which are more than 90%. The model constructed in this paper has a high fitting degree (R^2 = 0.972). In addition, the text test performs well in low-level text, and the prediction accuracy of high-level text is not enough.

Keywords: Rhetorical features · text readability · text difficulty · readability formula

1 Foreword

In recent years, the study of graded reading in Chinese field has gradually entered the vision of researchers [1]. The concept of graded reading is to match reading materials of different difficulty with different levels of ability readers [2, 3]. How to make a reasonable difficulty level classification of reading materials is the necessary and first item for graded reading research [4]. That is to say, the key problem of graded reading is the formulation of reasonable reading evaluation standards, and one of the important topics is readability [5]. Researchers often measure the difficulty of the text by readability to evaluate and

Fund project: This work is supported partially by the General Project of the State Language Commission *Research on Intelligent Identification of Chinese Rhetoric Ability for Compulsory Education* (YB 145-56), the General Project of the 14th Five-Year Scientific Research Plan of the National language commission (YB145-16) and China Post Doctoral Science Foundation (2022M722231).

W. Yu and L. Xuan (Eds.): MLICOM 2023, LNICST 518, pp. 222–235, 2024.
https://doi.org/10.1007/978-3-031-71716-1_17

grade the reading material [6]. It refers to the degree or attribute to which a text is easy to read and understand [7]. The readability calculation formula constructed by the concept of readability is often used in the study of reading grading, which becomes the reference basis for reading ability evaluation and the selection of reading materials [8].

The most important factor affecting text readability comes from the language itself. Therefore, researchers mostly analyze and elaborate the influencing factors of text readability from the language level of Chinese characters, words, sentences, and chapter features. This paper explores the influence of rhetorical features on text readability, and then constructs a calculation formula model of text readability combined with rhetorical features. The main research ideas include: (1) collecting the existing graded text corpus, marking out rhetorical features, and constructing the corpus with rhetoric annotation; (2) counting the values of rhetorical features, calculating the correlation coefficient to the difficulty of the rhetoric indicators, analyzing the calculation results, discussing the correlation between rhetoric features and text readability; (3) selecting the features from Chinese characters, words and sentences, to construct the calculation formula model of text readability, and then verifying the effect.

2 Related Studies

2.1 Study on the Readability of Chinese Text

Text readability is determined by the reader's understanding ability to the text [9]. The reader's understanding ability to the text, or the readability of the text, is actually influenced by many factors. From the external conditions, reading text format, typesetting, reading environment, media and other factors will affect the text readability. Form the internal conditions, Chinese characters, words, sentences, chapters directly determine the content quality of the text, which are the core factors of text readability.

Barman [10] mentioned six important factors that affecting the text readability in his study, which all come from the text language characteristics themselves. Dale and Chall [11] mentioned the external factors that affecting text readability in his definition of readability. When defining readability, Harris [12] mentioned that readability is influenced by the writing style of the text. Based on this, researchers try to find out the relevant factors affecting readability, comprehensively analyze the interaction of a series of influencing factors, so as to construct the readability formula, as an important means to measure the text readability. The first readability formula Come out in the United States in 1923 [13]. In 1928, Vogel [14] used multivariate linear regression equation to analyze the text characteristics, and constructed the earliest readability formula based on multivariate linear regression equation, which has a profound influence on the later study of readability formula.

Table 1 shows some English readability formulas developed by foreign researchers:

Chinese text readability studies date back to the 1970s and are mainly used to evaluate text difficulty. Yang Xiaoying [15] used the factor analysis method to analyze and summarize the key factors that affecting the difficulty of Chinese text, and selected the number of words, the number of sentences, the average number of strokes and other characteristic indicators to construct the earliest Chinese readability formula. Jing Xiyu

Table 1. Some English readability formulas

Formula name	computational formula	Adopt indicators
Flesch Reading Ease (Flesch, 1948)	Reading ease = 206.876-(1.015 Average sentence length) - (84.6 average number of syllables)	Length and syllable number
New Reading Ease (Flesch, 1951)	Reading ease = 1.599 Monosyllabic word ratio per 100 words-1.015 average number of words per sentence-31.517	Number of monosyllables and words
Flesch Grade Level (Kincaid et al., 1975)	Grade Level = -15.59 + (0.39 average sentence length) + (11.8 average number of syllables)	Length and syllable number
The New Dale-Chall (Chall & Dale, 1995)	$\frac{difficultwords}{totalwords}$ Grade Level = (0.1579) + (0.0496 average sentence length) + 3.6365	Ratio of difficult words, sentence length

[16] took the Chinese textbook (Taiwan textbook) as the basic corpus, used the number of strokes, sentence length, and the ratio of common words and characters as the parameters, the readability formula with grade as the dependent variable is constructed to predict the difficulty level of the text. Sun Gang [17] analyzed the surface features, part of speech features, syntactic tree features and information entropy features of Chinese text, and used 76 indicators from these four levels to predict the difficulty of Chinese text. Cheng Yong [18] based on the corpus of Chinese textbooks, analyzed the affecting to the text difficulty of 52 factors on Chinese characters, words, sentences and chapters level, and seven items which had a great influence on the text difficulty were included in the readability formula, that is, word frequency, word meaning richness, proportion of conjunctions, proportion of object word, proportion of action word, proportion of associated word and variation of sentence length. Wu Siyuan [19] began from the four levels of Chinese characters, words, syntax and chapters, constructed a Chinese text difficulty prediction feature system with 104 indicators in 13 dimensions. Li wenbiao [20] used the neural network model into Chinese text difficulty classification experiment, constructed the Chinese text difficulty classification model both in the traditional machine learning and mainstream neural network model method, at the same time they built an artificial Chinese text classification evaluation corpus, and the experiment achieved an ideal result, this model performance is better than the traditional method. In addition, the study on the readability of Chinese text can also carry out the construction of readability formulas in specific fields, such as the study of primary school Chinese textbooks with primary school children as the research object [21], and the research on the readability formula of Chinese as a foreign language to Chinese language learners [22, 23].

Table 2 are some of the Chinese readability formulas developed by the researchers:

Table 2. Some Chinese readability formulas

Formula name	computational formula	Adopt indicators
Yang Xiaoying (1970)	Grade = 0.1788 strokes over 10 strokes percentage + 0.1432 Average sentence length + 0.6375 percentage of difficult words Semester = 14.95961 + 39.07746 number of words-2.48491 average number of strokes + 1.11506 number of sentences	Stroke, difficult word ratio, sentence length, number of words, number of sentences, number of strokes
Jing Xiyu (1995)	Grade = 8.76105604 +.00272438 Text length + 0.07866782 Average sentence length-8.9311010 Common word ratio + 0.42920182 poetry style + 3.23677141 classical Chinese style	Text length, sentence length, common word ratio, style
Cheng Yong (2020)	Difficulty level = 38.36–45.65 average word frequency + 54.92 proportion of conjunctions-8. 96 proportion of object words + 11.13 word meaning richness-12.34 proportion of action words + 0. 012 variation of sentence length + 20 proportion of associated words	Word frequency, word meaning richness, proportion of conjunctions, proportion of object words, proportion of action words, proportion of associated words and variation of sentence length

2.2 Research on Rhetoric and Rhetoric Teaching

Rhetoric is the process of selecting language materials [24], is a variation of grammatical phenomena [25], and is interrelated with human cognitive thinking [26] and aesthetic ideas [27]. The essence and classification of rhetoric, the analysis of speech style and the theory of rhetoric structure are important topics in the study of Chinese rhetoric. On a macro level, rhetoric research is a comprehensive discussion of the law to the language communication; On a micro level, rhetoric research focuses on the choice and organization of language structure [28]. For rhetoric teaching, it is necessary to focus on the rhetoric means to choose the appropriate language structure to improve the effect of language expression in a specific context, including phonetic rhetoric, word rhetoric, sentence rhetoric, text rhetoric, rhetoric figure and language style, etc.

In this regard, Phonetic rhetoric reflects the symbolic relationship between speech and meaning [29], such as the use of onomatopoeia. Word rhetoric refers to the use of words that should differentiate the meaning of words, clear the reference, exact the number, distinguish praise and criticism, see the structure, change the shape of words and etc. [30]. Sentence rhetoric pays attention to the refined sentence and its connection with context, style and figure, and the relationship between different sentences [31]. The

text rhetoric focuses on the text structure, text cohesion and coherence, text information and context [32]. Rhetoric figure is the figure of speech. It is the most distinctive content in rhetoric teaching. In the stage of basic education, there are eight common rhetorical devices: metaphor, personification, exaggeration, parallel, antithesis, repetition, question and answer and ask in reply.[1] Speech style reflects the functional style of the language [33]. It was divided into spoken style and written style, or elegant style and vernacular style [34]. Written style can be further divided into document style, language style, scientific language, political style and literary style.

The purpose of rhetoric teaching is to cultivate students' language expression ability, and to enhance and expand the ability of aesthetic appreciation and logical thinking ability [35]. In the stage of basic education, the identification and application of rhetoric figures is a more prominent and common rhetoric teaching content. For example, metaphor, the most important rhetorical figures is the one that has the largest proportion and is the most widely used in rhetoric teaching [36]. The teaching requirements of metaphor rhetoric also develop with different teaching stages: the primary school only needs to simply understand the use of metaphors. But in the junior high school, they should further learn metaphor rhetoric on the basis of practical application and appreciation evaluation.

3 A Graded Rhetoric Labeled Chinese Text Corpus

In this paper, A graded Rhetoric labeled Chinese text corpus was constructed. It is classified into 11 difficulty levels and marked with 15 rhetorical devices. The difficulty classification criteria and rhetoric annotation information are described as follows.

3.1 Difficulty Graded Data

The corpus of this paper comes from the corpus of Chinese textbooks collected from the public website,[2] including the text data of all grades from primary, junior and high school stage textbooks of the ministry compiled edition. After filtering out some special genre texts, such as poetry, drama and classical Chinese, we constructed a corpus of Chinese textbooks for primary and secondary schools, including 250 primary school textbooks, 118 junior high school textbooks and 104 senior high school textbooks.

3.2 Rhetoric Annotation

Rhetoric phenomenon generally exists in Chinese text, and it also attaches great importance to rhetoric teaching in Chinese curriculum. Chen Wangdao pointed out in *The Origin of rhetoric* that there are two differences between positive rhetoric and negative rhetoric, among which the study of positive rhetoric phenomenon focuses on the division of rhetoric pattern. The classification and definition of rhetoric has always been an important topic of discussion in the rhetoric circle, which is of guiding significance to rhetoric teaching. It is mentioned in *Explanation of grammar and rhetorical knowledge*

[1] http://www.moe.gov.cn/srcsite/A26/s8001/202204/t20220420_619921.html.
[2] http://www.dzkbw.com/.

from annex 3 of the Chinese Curriculum Standard for Compulsory Education (2022 Edition) that there are 8 common rhetorical devices: metaphor, personification, exaggeration, parallelism, antithesis, repetition, question and answer and ask in reply. In addition to the 8 common rhetorical devices mentioned in the curriculum standards, this paper finds out 7 kinds of uncommon rhetorical devices according to the content of primary and secondary school texts, and refers to the connotation of the corresponding rhetorical figure in *The Origin of rhetoric* to design a rhetoric device annotation scheme.

4 Correlation Between Rhetorical Features and Text Readability

4.1 Characteristic Interpretation

Rhetoric is an important means of language expression, and the rhetorical characteristics of a text are highly correlated with the text readability. This paper analyzes the characteristics of rhetoric based on the corpus of rhetoric annotation, including The number of rhetoric device used sentences, rhetoric device sentence usage, the number of rhetorical device species, the proportion of common rhetoric device and the number of various kinds of rhetoric, among which:

(1) The number of rhetoric device used sentences refers to the number of all sentences that use rhetoric device in a certain level of text.
(2) rhetoric device sentence usage represents the ratio of the number of all sentences using rhetoric in a certain level of text.
(3) The number of rhetorical device species represents the number of all rhetoric used in a certain level of text.
(4) The proportion of common rhetoric device refers to the proportion of the total number of metaphors, personification, exaggeration, parallel, dual, repetition, asking and rhetorical questions to the total number of rhetoric used in a certain level of text.
(5) the number of various kinds of rhetoric device is the number of 15 kinds of devices used in every level of text: metaphor, personification, exaggerate, parallelism, antithesis, repeatedly, question and answer, ask in reply, metonymy, a word or phrase with double meaning, quote, parable, centre, synaesthesia, irony.

4.2 Correlation Calculation

This paper analyzes the correlation of rhetorical features and text readability based on the rhetoric annotation corpus. The main steps are described as follows:

(1) Count: Count and calculate the above rhetorical feature data, and normalize the statistical results.
(2) Correlation calculation: calculate the difficulty correlation of the statistical results, that is, take the grade as the text readability level standard to calculate the text readability correlation coefficient of each rhetorical feature. The correlation coefficient formula is such as Formula 1, where x represents the text readability level, y represents the statistical results of each rhetorical feature data, and i represents the i^{th} rhetorical feature.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

4.3 Results and Analysis

According to statistics:

(1) The number of rhetoric device used sentences, number of rhetorical device species, metaphor, exaggeration, parallelism, question and answer, ask in reply, metonymy, a word or phrase with double meaning, quote and parable show an upward trend with the increase of the difficulty level. The number of some rhetoric device in low-level textbooks is significantly less than that of high-level textbooks, and even the occurrence is 0.
(2) The proportion of common rhetoric device and the usage of rhetoric generally show a downward trend with the increase of the difficulty level. The proportion of first grade textbooks reached 97.2%, and the usage rate was 49.6%, while the proportion of second grade textbooks was 79.6% and only 8.1%. In addition, the number of anthropomorphic rhetoric used in the low-level text is very high, and the higher the level is, the number of its uses tends to decline.
(3) The first three indicators with the highest correlation with the difficulty of the text are the number of quote device, ask in reply device and the proportion of common rhetoric device.

5 Text Readability Calculation Formula

As can be seen from the above, text readability has a strong correlation with some rhetorical features. In this paper, we combine rhetorical features and other language characteristics to construct the calculation formula of text readability to evaluate the level of text difficulty. The specific steps are described as follows:

(1) Index selection: According to the calculation results of correlation, set a threshold value, and the difficulty correlation coefficient is within the threshold index, and remove the index with little correlation.
(2) Model construction: Taking the selected characteristic measurement index as the independent variable and the difficulty level as the dependent variable, the text readability calculation formula model is constructed through multiple linear regression equation analysis.
(3) Effect verification: the constructed text readability calculation formula model is tested for goodness of fit, and a certain number of examples are selected into the model for text test.

5.1 Index Selection

In addition to rhetorical features, this paper also examines the relevance of the features of Chinese characters, vocabulary and sentences to text readability among:

Chinese Character Level. Chinese characters are the recording symbol of Chinese text, and their internal structure is twists and turns. The reading memorization and understanding of Chinese characters are closely related to their form, quantity and usage frequency. Therefore, this paper analyzes the characteristics of Chinese characters by using the average number of strokes, the number of characters, the number of characters and the proportion of difficult characters.

Word Level. Word is the building material of language, and is the brick of text. The word usage is an important factor to measure the readability of the text, which is mainly reflected in the number of word, word attributes and the frequency of word use. In This paper, the lexical level characteristics of the text are analyzed by using such indicators as the number of words, the number of word styles, part of speech ratio, average word frequency and low frequency word ratio.

Sentence Level. Sentence is the most basic unit of language expression and a very important variable in the analysis of text readability. The number and length of sentences, the richness of sentence classes, and the structural complexity of sentences may all affect the assessment of text readability. This paper analyzes the sentence level characteristics of the text by sentence number, average sentence length, single sentence proportion, statement sentence proportion and average clause number.

Combined with the statistical results of rhetorical features, the difficulty correlation coefficient of feature-related indicators at all levels is shown in Table 3:

Table 3. Difficulty coefficient of relevant indicators at all levels

Language level	characteristic index	degree of relevance	Language level	characteristic index	degree of relevance
Chinese character level	The number of character styles	0.911	Rhetoric level	The number of rhetoric device used sentences	0.620
	character number	0.956		Rhetoric device sentence usage	−0.693
	Difficult character ratio	0.958		Number of rhetorical device species	0.821
	The average number of strokes	−0.626		The proportion of common rhetoric device	−0.911
Word level	The number of word styles	0.964		metaphor	0.850
	The number of words	0.939		personification	−0.604
	Average word frequency	0.565		exaggerate	0.563

(*continued*)

Table 3. (*continued*)

Language level	characteristic index	degree of relevance	Language level	characteristic index	degree of relevance
	The proportion of low frequency words	0.632		parallelism	0.413
	The proportion of nouns	−0.560		antithesis	0.082
	The proportion of adjectives	−0.762		repeatedly	0.197
	The proportion of Verbs	−0.681		question and answer	0.933
	The proportion of adverbs	0.808		ask in reply	0.944
	The proportion of prepositions	0.514		metonymy	0.752
	The proportion of pronouns	0.808		a word or phrase with double meaning	0.569
Sentence level	Number of sentences	0.954		quote	0.945
	Average sentence length	0.943		parable	−0.972
	The average number of clauses	0.931		centre	0.399
	The proportion of simple sentences	−0.101		synaesthesia	−0.219
	The proportion of statements	−0.264		irony	0.197

Combined with the value of the relevant characteristic indicators at the rhetoric level, the correlation coefficient threshold was set to 0.4, and 31 indicators with the absolute value of the correlation coefficient retained greater than 0.4 were selected, including 4 Chinese character level indicators, 10 vocabulary level indicators, 3 sentence level indicators and 14 rhetoric level indicators. As shown in Table 4.

Table 4. The retained 31 indicators

Chinese character level	The number of character styles, character number, Difficult character ratio, The average number of strokes
Word level	The number of word styles, The number of words, Average word frequency, The proportion of low frequency words, The proportion of nouns, The proportion of adjectives, The proportion of Verbs, The proportion of adverbs, The proportion of prepositions, The proportion of pronouns
Sentence level	Number of sentences, Average sentence length, The average number of clauses
Rhetoric level	The number of rhetoric device used sentences, Rhetoric device sentence usage, Number of rhetorical device species, The proportion of common rhetoric device, Metaphor, personification, exaggerate, parallelism, question and answer, ask in reply, metonymy, a word or phrase with double meaning, quote, parable

5.2 Model Construction

In order to construct the readability formula, this paper first transforms the difficulty level to grade number according to the age; then use the 31 index values selected in the previous paper as independent variables, and the difficulty level as the dependent variable into multiple linear regression equation to construct the readability formula model. The analysis results are shown in Table 5.

Table 5. Results of the multiple linear regression analysis

Model		Unstandardized coefficients		Standardization coefficient	t	conspicuousness
		B	standard error	Beta		
1	(constant)	14.303	3.749	–	3.815	0.009
2	The number of word styles	0.020	0.562	0.444	3.651	0.011
3	Difficult character ratio	47.973	2.168	0.934	4.887	0.003
4	The number of characters	−0.003	0.001	−0.558	−3.459	0.013
5	The proportion of common rhetoric device	−8.907	2.684	−0.118	−2.418	0.010

The readability formula model constructed based on the table content is:

$$\begin{aligned}&\text{Difficulty level} = 14.303 + 0.02\\&\text{The number of word styles} + 47.973\\&\text{Difficult character ratio} - 0.003\\&\text{The number of characters} - 8.907\\&\text{The proportion of common rhetoric device}\end{aligned} \tag{2}$$

5.3 Effect Verification

1. The Degree of Fit Calculation

For the readability formula model, the goodness of fit test is conducted, and the calculation method is as follows:

Let y be the value to be fitted, the mean is and the fitted value are. Note: $\overline{y}\hat{y}$

$$\text{Sum of Squares Total (SST)} : \sum\nolimits_{i=1}^{n} (y_i - \overline{y})^2 \tag{3}$$

$$\text{Regression Sum of squares (SSR)} : \sum_{i=1}^{n} \left(\hat{y}_i - \overline{y}\right)^2 \tag{4}$$

$$\text{Error Sum of Squares (SSE)} : \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{5}$$

Then we have: SST = SSR + SSE coefficient of determination:

$$R^2 = \frac{SSR}{SST} \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} == 1\text{-}\frac{SSE}{SST} \tag{6}$$

The final model fit $R^2 = 0.972$.

2. Text Testing

In terms of text testing, this paper selects examples for verification and analysis. The verification process is as follows:

(1) Select 5 articles from each level of text and mark the text difficulty level;
(2) Using computer technology to count the number of words, difficult words, the number of words and common rhetoric of each article;
(3) The four parameter values are inserted into the readability formula to calculate the difficulty level and compare it with the annotated level.

For example, in the third grade article “Primary School under the Big Green Tree”, the text difficulty level is marked as 8, the number of words is 89, the number of words is 177, the proportion of difficult words is 0.028, the proportion of common rhetoric is 1, and the statistical results into the readability formula to calculate the difficulty level

Table 6. Text difficulty level verification results at every level

Test group 1	6	8	8	8	9	10	19	19	18	13	9
Test group 2	6	6	7	8	10	13	10	16	15	14	23
Test group 3	6	7	8	16	10	10	12	15	16	12	24
Test group 4	9	9	8	10	8	18	13	19	11	16	30
Test group 5	7	7	7	7	15	13	14	15	21	21	29
Average level	6.8	7.4	7.6	9.8	10.4	12.8	13.6	16.8	16.2	15.2	23
The annotation level	6	7	8	9	10	11	12	13	14	15	16
error	0.8	0.4	−0.4	0.8	0.4	1.8	1.6	3.8	2.2	0.2	7

is about 8. Thus, the difficulty level of the article calculated by the readability formula is consistent with the actual text difficulty level.

The verification results of the test levels are shown in Table 6:

According to the verification results, in the lower grade text, the error between the average level and the annotation level calculated by the formula is small, and the prediction accuracy is high; the prediction effect of the middle and higher grade text is not as good as that of the lower grade, and the prediction value calculated by the formula is generally higher than the marked true value, and the prediction accuracy is low. It can be seen that the model is unstable in the text with higher difficulty level, and it is more suitable for the evaluation and measurement of text difficulty with lower difficulty level.

6 Summary

This paper studies the readability of Chinese text and analyzes the relevant characteristics of the text readability from the perspective of rhetoric and rhetoric teaching research. It constructs a graded corpus of Chinese text which annotated with 15 kinds of rhetoric. Based on the corpus, this paper get some statistical analysis, and calculate the correlation between rhetoric features and text readability. It shows that there are highly correlation between the two items. As it is known that text readability is actually affected by many factors. This paper further selects relevant indexes combined with rhetoric, Chinese character, word and sentence level features, and jointly constructs a new formula model for text readability calculation. The model can objectively and regularly divide the difficulty level for reading text, and has the characteristics of simple and easy to operate.

References

1. Zhang, W., Zhang, K.: The dilemma and breakthrough of Chinese intelligent reading. Lang. Strategy Res. **3**(04), 70–77 (2018)
2. Zhou, J., Zhang, W., Zhang, K., et al.: Construction of machine intelligence aided reading quantitative model based on logical image theory. Lang. Appl. **03**, 96–104 (2019)
3. Wang, Q.: Principles and Countermeasures of Graded Reading. The Young Children Study **2011**(02), 4–7 + 27 (2011)

4. Xu, M.: Development and description of the text difficulty grading criteria. Curriculum Teach. Res. Shanghai **01**, 49–56 (2020)
5. Song, Y., Chen, R., Li, Y., et al.: Discussion on the readability of Chinese text: index selection, model establishment and validity validation. China J. Psychol. **1**, 75–106 (2013)
6. Wu, S., Cai, J., Yu, D., Jiang, X.: Review of automated analytical studies on text readability. Chin. Inform. J. **12**, 1–10 (2018)
7. Zhou, D., Zheng, Z.: Review of readability studies. J. Quanzhou Normal Univ. **38**(01), 55–63 (2020)
8. Wang, L.: Research on the development stage and characteristics of the text readability formula. Lang. Teach. Res. **02**, 29–40 (2022)
9. Klare, G.R.: Readability. In: Pearson, P.D., Barr, R., Kamil, M.L., Mosenthal, P. (eds.) Handbook of Reading Research, pp.681–744. Longman, New York, London (1984)
10. Brabham, E.G., Villaume, S.K.: Leveled text: the good news and the bad news. The Reading Teacher (55), 438–441 (2002)
11. Chall, J.S., Dale, E.: Manual for the new Dale-Chall readability formula. Brookline, Cambridge (1995)
12. Harris, A.J., Jacobson, M.D.: A framework for readability research: moving beyond Herbert Spencer. J. Read. **22**, 390–398 (1979)
13. Lively, B.A., Pressey, S.L.: A method for measuring the "vocabulary burden" of textbooks. Educ. Administ. Supervis. **9**, 389–398 (1923)
14. Vogel, M., Washburne, C.: An objective method of determining grade placement of children's readingmaterial. Element. School J. **5**, 373–381 (1928)
15. Yang, X.: Practical Chinese newspaper readability formula. Res. Journalism **13**, 37–62 (1974)
16. Jing, X.: Study on the suitability of Chinese textbooks: the estimation of the grade value of Chinese. Educ. Res. Inf. **3**, 113–127 (1995)
17. Sun, G.: Research on the readability prediction method of Chinese text based on linear regression. Nanjing University (Master) (2015)
18. Cheng, Y., Xu, D., Dong, J.: Analysis on the key factors of text reading difficulty classification based on the corpus of Chinese textbook. Lang. Appl. **01**, 132–143 (2020)
19. Wu, S., Yu, D., Jiang, X.: Chinese text readability feature system construction and validity verification. World Chin. Teach. **01**, 81–97 (2020)
20. Li, W., Wu, Y.: The Chinese text difficulty classification based on the neural network model. Chin. Inf. J. **37**(02), 158–168 (2023)
21. Liu, M., Jiang, Y., Li. Y., Li. H.: Research and application of Chinese readability formula based on primary school Chinese textbooks. In: Chinese Psychological Society. Summary Set of the 21st National Academic Conference on Psychology. Chinese Psychological Association: Chinese Psychological Association (2018)
22. Wang, L.: Study on the readability of Chinese text among junior and junior Japanese and Korean learners. Lang. Teach. Res. **05**, 15–25 (2017)
23. Zuo, H., Zhu, Y.: Study on Chinese text readability formula for intermediate European and American students. World Chin. Teach. **02**, 263–276 (2014)
24. Chen, W.: Rhetoric is Sent to All. Fudan University Press, Shanghai (2008)
25. Lu, C.: New rhetorical argument. J. Zhejiang Norm. Univ. (Soc. Sci. Edn.) **1**, 82–89 (1985)
26. Zhang, Z.: Thoughts on the essence of rhetoric, namely, the identity of rhetoric. Rhetoric Study **3**, 29–30 (2002)
27. Zong, T., Chen, G.: Aesthetic care of Chinese rhetoric. J. Liaodong Univ. (Soc. Sci. Edn.) **4**, 92–96 (2017)
28. Zhang, Z.: Rhetoric is a Choice Process. Rhetoric and Rhetoric Teaching. Shanghai Education Press, Shanghai (1985)
29. Chang, Z.: Primary school Chinese rhetoric teaching under the perspective of unified teaching materials. Chin. Construct. **24**, 8–12 (2022)

30. Liu, H.: Word Words. Jiangxi People's Publishing House, Jiangxi (1980)
31. Ni, B.: Try to Find the Best Turn of Phrase. Shanghai Education Press, Shanghai (1985)
32. Zheng, G.: Chinese chapter rhetoric. Foreign Languages Press, Beijing (2002)
33. Tao, H.Y.: Try the grammatical meaning of the classification of the Analects of Confucius. Contemp. Linguist. (03), 15–24 + 61(1999)
34. Feng, S.: The mechanism of the Analects and its grammatical properties. Chin. Lang. **2010**(05), 400–412 + 479 (2010)
35. Yang, X.: Rhetoric Teaching in Junior High School Chinese Teaching. Management Office of National Teacher Research Fund. Collection of the First "Education teaching and Innovation Research" Forum in the New Period. [Unknown Publisher], pp. 86–88 (2022)
36. Lu, Y.: Research on the Teaching of Junior High School Chinese Metaphor. Guizhou Normal University (2023)