



Research into Reading Grants

ISSN 2755-2543 Report 1/2022



Applying the Lexile Framework to evaluate text complexity in the Chinese context: A corpus-based study

RESEARCH IN EFL READING AND READING ASSESSMENT

Jason Fan Deputy Director and Senior Research Fellow, Language Testing Research Centre University of Melbourne, Australia

Tan Jin Professor, School of Foreign Studies, South China Normal University, China

Ivy Chen

Research Fellow, Language Testing Research Centre, University of Melbourne, Australia

© BRITISH COUNCIL AND METAMETRICS 2022

ABSTRACT

The Lexile Framework is an innovative approach to reading comprehension whose validity has been proven through over 20 years of ongoing research. The use of the Lexile Framework has greatly facilitated the teaching, learning and assessment of reading, which is considered as a vitally essential language skill for success at all levels of education. Despite the extensive recognition of its validity and use in numerous contexts, the Lexile Framework is not widely known in China; nor are the Lexile measures familiar to English language educators, learners and assessment developers in China.

This study aimed to explore the extent to which the Lexile Framework was applicable to the Chinese educational context by analysing the Lexile measures of the reading texts in four prominent, high-stakes English tests in China, representing reading demands at four distinct educational stages: junior secondary, senior secondary, collegiate basic and collegiate intermediate. The four English tests are: the English component of the Senior High School Entrance Examination (SHSEE); the National Matriculation English Test (NMET); and the College English Test Band 4 (CET4) and Band 6 (CET6).

To further interrogate the accuracy of Lexile measures in capturing the text complexity of the test materials, we also compared the text complexity results generated by the Lexile Analyzer, which implements the Lexile Framework, and Eng-Editor, a locally developed text complexity evaluation tool.

Two research questions were investigated in this study: 1) What is the range of the Lexile measures for the reading texts in English test papers at each of the four educational levels in China? 2) In what ways are Lexile measures comparable to text complexity estimates generated by Eng-Editor? To address the two research questions, a total of 800 benchmark texts, with 200 at each level, were randomly selected from a corpus which consisted of reading texts from the four high-stakes English tests in focus. These benchmark texts were then analyzed with the Lexile Analyzer and Eng-Editor respectively, which generated text complexity measures across multiple dimensions for each text. Multiple statistical procedures were implemented, including descriptive statistics, ANOVA, correlation analysis, ordinal logistic regression analysis, and linear regression analysis.

The results indicate that the Lexile measures of the benchmark texts could effectively determine their classification into the four educational levels in focus; there was a strong relationship between the Lexile measures of the benchmark texts and the corresponding complexity estimates produced by Eng-Editor. The findings of this study provide empirical support to the applicability of the Lexile Framework to English language education in China. Meanwhile, through the application of the Lexile Framework, this study also sheds light on the English reading demands of China's four educational levels in the global context.

AUTHORS

Dr Jason Fan

Dr Jason Fan is the Deputy Director and Senior Research Fellow at the Language Testing Research Centre at the University of Melbourne. His research interests include language test validation and research methodology. He is the co-author of *Fairness, Justice and Language Assessment* (2019, OUP, with Tim McNamara and Ute Knoch), and *Working Towards a Proficiency Scale for Business English Writing: A Mixed-Methods Study* (2021, Springer, with Li Wang). He is the second Vice-President of the Asian Association for Language Assessment (AALA) (from 2022), and the co-editor (with John Read) of *Papers in Language Testing and Assessment* (from 2021).

Professor Tan Jin

Tan Jin is a Professor in the School of Foreign Studies at South China Normal University, China. His research interests include corpus linguistics, language assessment, as well as computer-assisted language learning. He has specific expertise in the development of computational tools for assessing text complexity, particularly for the selection and adaptation of reading materials. He and his team developed a data-driven text adaptation system between 2014 and 2022. In addition, he has constructed a corpus of expert essays, which he and his team are using to explore the linguistic realisations of rhetorical functions in narrative and argumentative writing.

Dr Ivy Chen

Dr Ivy Chen is a Research Fellow at the Language Testing Research Centre, University of Melbourne. Her research interests include language testing, SLA, and corpus linguistics. Her PhD dissertation, *A corpus-driven receptive test of collocational knowledge*, employed the argument-based approach to validation and modelled factors affecting item difficulty using multilevel path analysis. Ivy has lectured in applied linguistics and ESL at the tertiary level and has experience teaching ESL/EFL and test preparation in several countries and regions such as Australia, Taiwan and the US.

CONTENTS

1.	INTR	ODUCTION	6
2.	LITE	RATURE REVIEW	7
2.	1 Us	sing the Lexile Framework in measuring text complexity	7
2.	2 Er	glish language education and assessment in China	10
3.	THE	PRESENT STUDY	12
4.	METH	HODS AND MATERIALS	13
4.	1 The	e four steps in this study	13
4.	2 The	e corpus	14
4.	3 The	e benchmark texts	14
4.	4 Tex	xt complexity analysis	15
4.	5 Sta	atistical analysis	16
5.	RESU	JLTS	16
5.	1 Te	ext complexity results generated by the Lexile Analyzer	16
ļ	5.1.1	Descriptive statistics	16
Į	5.1.2	Comparing the Lexile measures at the four educational stages	20
Į	5.1.3	Using Lexile measure to predict educational level	21
5.	2 Co	omparing text complexity results of the Lexile Analyzer and Eng-Editor	23
!	5.2.1	Text complexity results generated by Eng-Editor	23
!	5.2.2	Comparing text complexity results of the Lexile Analyzer and Eng-Editor	26
6.	SUMI	MARY AND DISCUSSION	27
REF	EREN	ICES	31

List of tables

Table 1: A summary of the Lexile measures across different educational levels	9
Table 2: The Lexile measures of the reading texts used in Eiken tests	10
Table 3: Text samples used in the study	14
Table 4: Descriptive statistics of complexity measures generated by Lexile Analyzer across the four educational stages	18
Table 5: Median lexile measures at the four educational stages	20
Table 6: Mapping the complexity levels assigned by Eng-Editor and educational levels	24
Table 7: Correlation matrix between Eng-Editor complexity levels and	
educational levels (n = 800)	25

List of figures

Figure 1: The procedural diagram of this study	13
Figure 2: Text complexity distribution by educational level	19
Figure 3: Histogram of the Lexile measures at the four educational levels	20
Figure 4: Means of the Lexile measures at the four educational levels (with 95% CI)	21
Figure 5: Scatterplots of the Lexile measure and the probability of a benchmark text at each level	22
Figure 6: Correlations between educational level and text complexity level	25
Figure 7: Mapping Lexile measures with Eng-Editor text complexity levels	26

1. INTRODUCTION

The Lexile Framework is an innovative approach to reading comprehension whose validity has been proven through over 20 years of ongoing research (Lennon & Burdick, 2004). The Framework consists of two components: the Lexile measure and the Lexile scale. The Lexile measure provides a numeric representation of the difficulty of a text and the ability of a reader; the Lexile scale is a developmental scale which ranges from below 200L (beginning readers) to above 1700L (advanced readers) (Wilkins, Hartman, Howland, & Sharma, 2010). According to Stenner (1996), the psychometric model underpinning the Lexile Framework is the Rasch measurement model, a one-parameter conjoint model which enables the calibration of text difficulty and reading ability on the same metric scale. A clear advantage of using the Rasch model in developing the Lexile scale lies in its 'specific objectivity' which means 'an absence of a conditional relationship (interaction) between reader and text' (Stenner, 1996, p. 318). In other words, a text's readability does not need to be evaluated in reference to the readers' characteristics. The Lexile measures have been widely used by assessment developers and publishers. According to Lennon and Burdick (2004), more than 450 publishers indicate the Lexile measures in the books, newspaper and magazine articles that they publish. The use of the Lexile Framework has greatly facilitated the teaching, learning and assessment of reading, which is considered as a vitally essential language skill for success at all levels of education.

Despite the extensive application of the Lexile Framework and the multiple benefits of linking it to language teaching, learning and assessment materials (Lennon & Burdick, 2004), few attempts have been made to apply the Lexile Framework in the Chinese context. The only study that we found is Van Moere and Spiby (2018), which linked the Lexile measures with teaching and assessment materials used at English as a Medium of Instruction (EMI) schools in China. Though the findings of this study shed light on the application of the Lexile Framework to the Chinese educational context, learners in EMI schools account for only a tiny fraction of English language learners in China, and their English proficiency is not representative of China's EFL learner population. To date, the Lexile Framework has not been widely adopted in China; nor are the Lexile measures familiar to language educators, learners and assessment developers in the Chinese context.

This study explores the applicability of the Lexile Framework to the Chinese educational context by analysing the complexity of the reading texts in four high-stakes English language tests administered to students at four distinct educational stages in China, including: the English component of the Senior High School Entrance Examination (SHSEE, a.k.a. *Zhongkao*); the National Matriculation English Test (NMET, a.k.a. *Gaokao*); and the College English Test Band 4 (CET4) and Band 6 (CET6).

According to the China's Standards of English (CSE), the national language proficiency framework akin to the Common European Framework of Reference for Languages (CEFR), the difficulty levels of these four tests are located at four CSE levels (i.e., SHEEE – Level 3; NMET – Level 4; CET4 – Level 5; and CET-6 – Level 6; see also Jin, Wu, Alderson, & Song, 2017). The Lexile Framework, therefore, was expected to be sensitive enough to distinguish the complexity levels of the reading texts used in the four English tests, aligned to different CSE levels. (Note: The English version of the CSE is freely downloadable at: www.neea.edu.cn/xhtml1/report/18113/2797-1.htm.)

In this study, the Lexile Analyzer, which analyzes the text to assign it a Lexile text measure, was employed to map the text complexity trajectory across the four levels. To further interrogate the applicability of the Lexile Framework to the Chinese context, we also used Eng-Editor, a locally-developed text complexity evaluation tool (Jin & Lu, 2018; Jin, Lu & Ni, 2020), as an external criterion by comparing its text complexity analysis results with the Lexile measures generated by the Lexile Analyzer. (Note: Eng-Editor is freely available at: www.languagedata.net/tester/)

2. LITERATURE REVIEW

2.1 Using the Lexile Framework in measuring text complexity

The Lexile Framework is a common scale that links the English reading ability of individual readers and the complexity of reading texts, making it possible to match readers with texts of an appropriate complexity (i.e., high enough comprehension rate but difficult enough to be challenging). Lexile measures are used to numerically represent the two, while the Lexile scale is a developmental continuum that ranges from beginners to readers of advanced texts. Currently, Lexile measures are the most widely used measures for reading. For text complexity, they are typically calculated by combining two measures into an algebraic equation, the Lexile equation:

- a) semantic difficulty, operationalised as word frequency, and more specifically, as mean log word frequency in the corpus used by the Lexile Analyzer
- b) syntactic complexity, operationalised as sentence length (Lennon & Burdick, 2004).

Lexile measures are higher with higher semantic difficulty and with higher syntactic complexity. Similarly, the Lexile equation can be used for reading comprehension test items, the scores from which can then be used as a measure of reader ability, usually after adjustments using the Rasch model (Lennon & Burdick, 2004). Several studies have examined the reading demands of different education levels through Lexile measures. For instance, Williamson et al. (2015) quantified the text complexity of commonly used introductory textbooks in the UK by taking a roughly even number of texts from 10 top universities and compared the findings with already known Lexile measures for USA universities, showing that Lexile ranges were very similar across the two countries. In the English as a Foreign Language (EFL) context, Choi et al. (2012) reported that the reading demand of the English textbooks in South Korea at Grade 9 ranged from 510L to 850L.

The findings that emerged from some of these studies are summarised in Table 1 below, showing the ranges of Lexile measures for US Common Core State Standards, textbooks in English (the US, the UK, and at a Chinese EMI school), and English textbooks used in EFL contexts, against students' year levels. Overall, as indicated in this table, the Lexile measures of textbooks in English are higher than those of English textbooks.

Despite the burgeoning interest in measuring the complexity of English teaching materials with the Lexile measures, a paucity of research, however, has explored the complexity of language testing materials with the aid of the Lexile Analyzer. Bush, Koons, & Sanford-Moore (2018a) examined the English part of South Korea's College Scholastic Ability Test (CSAT), which aims to test readiness to study at university (1000L–1240L), while Negishi (2015) examined university entry tests of several Japanese universities (1030L–1370L). MetaMetrics (2016) examined the complexity of the reading texts used in the TEAP, an English test used for university admissions purposes (1150L–1210L) and the reading sections of EIKEN tests (see Table 2 for a comparison with CEFR levels). The Lexile measure ranges of the end-of-high-school/university entry tests seem to correspond to English textbook levels for Korea but seem to be mostly higher than those for Japan (see Table 1).

	Credel		Textbooks in English				English Textbooks	
	Year	US standards	US*	UK*	A Chinese EMI school		South Korea*	Japan
	1		230L–920L					
	2							
Flomontory	3	4201 10101		375L–855L				
Elementary	4	420L-1010L			-			
	5						90L-290L	
	6							
the standard state	7	925L-1185L	880L–1110L	845L–1055L			330L-770L/110L- 850L	
Junior nign	8							210L-480L
	9							
Conton binb	10			9201-11101	8001-1100-1			
Senior high	11			920L-1110L	000E-1100+L		870L-1250L	540L-950L
	12	1050L–1385L	920L-1220L					
University			1220L-1390L	1210L-1360L			~1250L-1400L**	850L-1060L

Table 1: A summary of the Lexile measures across different educational levels

Notes. * = Collapsed multiple year level groups' Lexile ranges (25th and 75th percentile) and reported the lowest Lexile measure of the lowest grade and highest of the highest grade. ** = Seoul National University (across faculties)

Summary of Lexile measures from Bush, Koons, and Sanford-Moore (2018b); Choi (2012); Koons, Bush, & Sanford-Moore (2018); Lee (2014); Negishi (2015); Supplemental information for Appendix A of the Common Core State Standards for English language arts and literacy: New research on text complexity. (n.d.); Van Moere and Spiby (2018); Williamson et al. (2012); and Williamson et al. (2015).

Test	Range	CEFR	Ministry of Education benchmarks	
Eiken Grade 1	1280L-1360L	C1		
Eiken Pre Grade 1	1180L-1200L	B2	English instructors	
Eiken Grade 2	1000L-1020L	B1	high appeal graduates	
Eiken Pre Grade 2	740L–780L	A2	nigh school graduates	
Eiken Grade 3	530L-570L		junior high school graduates	
Eiken Grade 4	460L-520L	A1		
Eiken Grade 5	300L-380L			

Table 2: The Lexile measures of the reading texts used in Eiken tests

Note. The Lexile ranges in this table are based on MetaMetrics (2016), and the CEFR comarison and benchmarks are based on EIKEN (2021).

2.2 English language education and assessment in China

English language education and assessment play a prominent role on China's educational landscape and in society at large. As China embraces economic reforms and the open-door policy, and becomes irreversibly interwoven into the world market, the significance of English cannot be overemphasised for students at all levels of study. Concomitantly, the government has formulated a host of policies to promote English language education throughout the country. For example, in the mid-1990s, English began to be taught from Grade 3 in primary education (Wang, 2008). The importance of English is also evidenced by the fact that the total number of English language learners in China exceeds that of English native speakers; and that English language education is a compulsory requirement for education at all levels in China, from primary to tertiary education (e.g., Jin, et al., 2017). Among the four language skills of listening, reading, writing and speaking, reading has traditionally been considered a critically important skill to Chinese English learners, thus occupying a significant position in the English teaching curriculum and assessment at all educational levels in China (Jin & Yang, 2018).

Given that this study focuses on English language tests in China, it is necessary to provide some background about English language testing in China's educational and social context. China has a long history and fine traditions in using testing and examination for selection purposes (e.g., Fan, 2019; Spolsky, 1995). Modern theory and practice in foreign language testing, however, did not emerge and thrive until the late 1970s, following the resumption of the National Matriculation Test (NMT), known in China as *Gaokao*. After about 40 years of development, China currently has numerous English language tests developed and implemented at national, municipal, and school levels, with tens of millions of test-takers every year.

English language tests developed and used in China are characterised by a few salient features. First, many of them are extremely large in scale and operation. The College English Test (CET) is a good case in point. Developed by a group of language teaching and testing experts back in the 1980s, the CET has an annual test population of around 20 million today, making it the largest of its kind worldwide (e.g., Fan & Frost, 2021; Jin & Yang, 2018). Second, many of these English tests are high-stakes because their results are often used to make important decisions which have significant ramifications for individual stakeholders (e.g., students, teachers, parents), institutions and society. For example, students' scores on the National Matriculation English Test (NMET), which is a compulsory subject in the *Gaokao*, are used to determine whether they can be admitted into a university, and by extension, the tier and type of the university, which could significantly affect their future career and life. It is not an exaggeration that being successful on the various English examinations is the key to success for many in China (Cheng, 2008). Finally, these English tests tend to have strong washback effects on English teaching and learning practices (e.g., Fan, 2018).

As mentioned previously, this study focused on four high-stakes English tests: the English component of SHSEE, NMET, CET4 and CET6. Of the four English tests, the SHSEE is the entry test that junior secondary school students need to pass in order to be admitted into senior secondary schools (Yan & Fan, 2021). The NMET represents one of the compulsory subjects in the College Entrance Examination whose results are used to determine whether students can be admitted into universities (Dong, Fan & Xu, 2021). The CET is a national English proficiency test targeting non-English major undergraduates in China which is administered at two levels: CET4 for those who have completed the foundation stage of College English (CE) learning, and CET6 for those who have completed the more advanced stage of CE learning (Zheng & Cheng, 2008). All four tests are high-stakes as success or failure on each of them has significant consequences for stakeholders, particularly students, relevant programmes and institutions, and society at large. Due to the high-stakes nature of these testing programmes, their development and administration follow rigorous quality control procedures.

According to China's National Curricula, six distinct stages of English language education can be identified: 1) elementary; 2) junior secondary; 3) senior secondary; 4) undergraduate (non-English major); 5) undergraduate (English major); and 6) graduate (non-English major) (Jin, et al., 2017). The CSE consists of nine proficiency levels corresponding to nine different educational stages in China (Ministry of Education, 2018). For example, Levels 1 to 3 correspond to elementary (Grade 1–3), elementary (Grade 4–6), and junior secondary respectively (Ministry of Education, 2018).

The four tests in this study are supposed to be taken at the end of four stages in China's education: 1) junior secondary (i.e., CSE Level 3); 2) senior secondary (i.e., CSE Level 4); 3) collegiate basic (i.e., CSE Level 5); and 4) collegiate intermediate (i.e., CSE Level 6), encompassing the majority of English language learners in China.

3. THE PRESENT STUDY

As previously mentioned, this project aimed to explore the extent to which the Lexile Framework was applicable to the Chinese educational context by analysing the complexity of the reading texts used in four high-stakes English language tests at four distinct educational stages in China. In addition, we also interrogated the applicability of the Lexile Analyzer by comparing its results with the text complexity measures generated by Eng-Editor, an online text complexity analysis tool developed by referring to local English curriculum standards (Guo, Chen, Lei, & Jin, 2021; Jin & Lu, 2018; Jin, Lu & Ni, 2020).

Though both the Lexile Analyzer and Eng-Editor take into account semantic or lexical difficulty and syntactic complexity in evaluating text complexity, the theories and working mechanisms that underpin the two analysis tools are different. As explained in Section 2.1, the Lexile Analyzer calculates text complexity by combining two measures: (a) semantic difficulty, operationalised as mean log word frequency in the corpus used by the Lexile Analyzer; and (b) syntactic complexity, operationalised as sentence length (Lennon & Burdick, 2004). Eng-Editor, however, operationalises lexical difficulty by referring to local English curriculum standards, which are in turn aligned to different CSE levels.

In China, the vocabulary required for compulsory English education has been specified in a curriculum vocabulary list (Jin, Li, & Li, 2016). According to the Eng-Editor user manual, those curriculum vocabulary lists, separated into five educational levels together with off-list words, are used to determine the vocabulary coverage of a text that is being analyzed in Eng-Editor. With regard to syntactic complexity, six measures are used: coordinate phrases per clause; complex nominals per clause; non-finite elements per clause; dependent clauses per clause; mean length of clause; and mean length of sentence (Jin, Lu, & Ni, 2020). Finally, an additional measure of total text length is included in the calculation of text complexity, which results in a level that reflects the corresponding educational level of the text in accordance with five out of nine levels in the CSE, from elementary school to university (i.e., Levels 3–7). Eng-Editor is an online tool which is freely available to registered users. According to its developer, the total number of registered users of the two tools on the website (Eng-Editor and Chi-Editor) exceeds 10,000 as of December 2021, most of whom are reading teachers and researchers. While the exact number of active users of Eng-Editor is unavailable, it is estimated to be around 5,000.

The following two research questions were investigated in this study:

RQ1. What is the range of the Lexile measures for the reading texts in the English subject of SHSEE, NMET, CET4 and CET6?

RQ2. In what ways are Lexile measures comparable to text complexity estimates generated by Eng-Editor?

The findings of this study have several important implications. First, since the difficulties of the four reading tests in focus are aligned to Levels 3 to 6 on the CSE respectively, the ranges of the Lexile measures of the reading texts in the four tests can provide credible evidence as to the effectiveness of the Lexile Analyzer in measuring the complexity levels of the test materials used in the Chinese educational context. Second, by comparing the results based on the Lexile Framework, a global framework for evaluating text complexity, and Eng-Editor, a locally-developed text complexity tool, the findings will shed light on the applicability of the Lexile Framework to the Chinese EFL teaching, learning and assessment context. Finally, the findings of this study are instrumental in situating the reading demands of China's different educational stages in the global context.

4. METHODS AND MATERIALS

4.1 The four steps in this study

This study followed four steps (see Figure 1). During the first step, we selected 800 texts from a corpus which consists of a unique collection of around 7,000 text samples from either textbooks or reading passages from high-stakes English tests, representing a range of Chinese EFL learners' English language proficiency levels, from elementary to collegiate. The 800 texts selected for this study were from the English test materials used at four levels from junior secondary to collegiate intermediate.

Next, we analyzed the complexity of the 800 texts with the Lexile Analyzer, following the analysis procedures laid out in the User Guide. (Note: The benchmark texts in this study were analyzed in the Professional Lexile Analyzer. We would like to extend our heartfelt appreciation to Nathan Bond from MetaMetrics for his generous support in our data analysis.)

During the third step, we performed statistical analysis of the analysis results (see Section 4.5 for details) which enabled us to map the text complexity trajectory across the four educational stages in focus. Finally, we compared the text complexity results generated by the Lexile Analyzer and Eng-Editor.





4.2 The corpus

The corpus used in this study was constructed in 2018. As mentioned previously, the corpus includes a unique collection of around 7,000 text samples representing a range of Chinese EFL learners' English language proficiency levels. The text samples were systematically collected from two types of sources, that is, English textbooks and the test papers of several high-stakes English language tests in China. On the one hand, about half of the texts in the corpus were collected from 23 packages of English textbooks published by 11 publishers. These textbooks were approved by the Chinese Ministry of Education for use as English language teaching materials in public schools at three educational stages within 12 grade levels: elementary school (Grades 1–6; around 2,100 texts), junior high school (Grades 7–9; around 700 texts), and senior high school (Grades 10–12; around 700 texts). The other half of the texts were collected from the test papers of four large-scale and high-stakes English proficiency tests in China, including the English component of the SHSEE (around 1,700 texts), the NMET (around 1,100 texts), the CET4 and CET6 (around 500 texts), and the Graduate School Entrance English Examination GSEEE (a.k.a. *Kaoyan*; around 200 texts). All the texts analyzed in this study were selected from the reading sections of the English components of *Zhongkao* and *Gaokao*, CET4 and CET6.

4.3 The benchmark texts

As mentioned above, we selected 800 benchmark texts from the four high-stakes English tests at four educational levels in the corpus, namely the English component of the SHSEE (junior secondary), the NMET (senior secondary), CET4 (collegiate basic), and CET6 (collegiate intermediate). To ensure the representativeness of the benchmark text samples, we randomly selected 200 texts from the sub-corpus at each level. Table 3 delineates the number and source of the benchmark texts selected for this study at each educational stage. After the texts were selected, we manually checked each of the 800 texts included in this study and made edits if necessary, following the text preparation guidelines and formatting conventions detailed in the Lexile Analyzer User Guide.

Educational stage		n	Test
1.	Junior secondary	200	The English component of SHSEE
2.	Senior secondary	200	NMET
3.	Collegiate basic	200	CET4
4.	Collegiate intermediate	200	CET6

Table 3: Text samples used in the study

Note: We included the text samples only from the four English tests without the questions that are associated with these texts because according to the Lexile Analyzer User Guide, inaccurate Lexile measures could be generated when the test items are analyzed together with the texts.

4.4 Text complexity analysis

As illustrated in Figure 1, the second step of this study involved the analysis of the complexity of the selected texts in both *Lexile*[®] Text Analyzer and Eng-Editor. It should be noted that the authors of this report, who are not trained MetaMetrics[®] staff, prepared the texts for submission to Lexile Text Analyzer. The results generated by the Lexile Analyzer include the following measures.

- Mean Log Word Frequency: This value is the logarithm of the number of times a word appears in the MetaMetrics corpus of over 1.4 billion words. According to the Lexile Analyzer User Guide, the value should be understood as how often a word occurs per 5 million words. The mean log word frequency is the average of all such values for words which appear in the text being analyzed.
- 2. **Mean Sentence Length**: This value is the average length of a sentence in the text, based on the sentences that were analyzed.
- 3. Lexile Measure: This value indicates the reading demand of the text in terms of the semantic difficulty and syntactic complexity. The Lexile scale generally ranges from below 0L for beginning reader texts to above 1600L for advanced texts.

To compare the analysis results from the two tools, we also analyzed the complexity of the selected texts with Eng-Editor which provides three complexity indices: lexical complexity, syntactic complexity and text complexity. Lexical complexity is calculated by comparing a text against five vocabulary lists from the Chinese EFL curriculum standards (one from the *English Curriculum Standards for Compulsory Education*, one from the *English Curriculum Standards for High School*, two from the *National College English Test Syllabus*, and one from the *College English Curriculum Requirements*) to check what percentage of the items in the text are covered by each list. Syntactic complexity is calculated based on six indices (see Section 3). The last index, text complexity, indicates the overall complexity of a text, and is calculated based on its lexical complexity, syntactic complexity, and text length. Each complexity estimate consists of two parts: (a) the level on the China's Standards of English (CSE); and (b) the percentile of the text in the corpus built by the developer in light of the three indices, that is, lexical difficulty, syntactic complexity and overall text complexity.

The user manual of Eng-Editor provides an example to assist its users in interpreting the analysis results. For example, the estimate of 6.61 for text complexity means that this text is located at Level 6 on the CSE (the '6' before the decimal), which corresponds to the collegiate intermediate level; the two digits after the decimal (i.e. '.61') indicates that in terms of overall text complexity, this text is approximately 61 percentile in the benchmark texts in the corpus at CSE Level 6. In other words, the complexity estimates generated by Eng-Editor provides both the complexity level of a text in accordance with the CSE level and its standing in comparison with the other texts in the developer's corpus at the same CSE level.

4.5 Statistical analysis

To address RQ1, we calculated the descriptive statistics of the Lexile measures of the benchmark texts at each educational stage, including mean, standard deviation, median, and interquartile range. The analysis results enabled us to map the text complexity distributions at and across the four educational stages in focus. In addition, we used one-way ANOVA to compare whether the Lexile measures at the four educational stages were significantly different. Finally, we used ordinal logistic regression analysis to investigate the extent to which the Lexile measure of a text could predict its classification into one of four educational levels from which the benchmark texts were selected.

To address RQ2, we mapped the complexity level of each text assigned by Eng-Editor with the educational level from which it was selected across the three dimensions of lexical complexity, syntactic complexity, and text complexity. Next, we calculated the classification accuracy. This analysis helped us gain an understanding of the general accuracy or dependability of Eng-Editor as an external criterion in this study in classifying a text into its corresponding CSE and/or educational level. To compare the complexity results generated by the two tools, we utilised Spearman correlation analysis. Finally, we also performed linear regression analysis to explore whether the Lexile measure of a text could predict its text complexity value provided by Eng-Editor. All statistical analyzes in this study were implemented in Minitab (Minitab, 2021).

5. RESULTS

In this section, we present the Lexile measures of the benchmark texts across the four educational stages in focus and the text complexity levels assigned by Eng-Editor before comparing the results generated by these two analysis tools.

5.1 Text complexity results generated by the Lexile Analyzer

5.1.1 Descriptive statistics

Table 4 shows the descriptive statistics of the text complexity measures generated by the Lexile Analyzer. The first measure, Mean Log Word Frequency, indicates how often a word occurs per 5 million words in the Lexile corpus. This value is calculated through averaging such values for all words in a text that is being analyzed. As indicated in Table 4, the mean value of this measure decreases as the educational stage progresses from junior secondary (M = 3.66, SD = 0.17) to collegiate intermediate (M = 3.30, SD = 0.12), signalling lower frequency and thus increasing difficulty of vocabulary in the test materials at a higher educational level.

The second measure, Mean Sentence Length, represents the average length of a sentence in a text that is being analyzed. Table 4 indicates that this measure increases as the educational stage progresses, suggesting that the sentences become longer and more complex in the test materials at a higher educational level. It is worth noting that this measure increases more noticeably from junior secondary (M = 12.08, SD = 2.71) to senior secondary (M = 16.22, SD = 3.91) and then to the collegiate basic level (M = 19.19, SD = 3.43); however, only a slight increase is observed from collegiate basic to collegiate intermediate (M = 19.87, SD = 3.87).

Word Count is the third measure provided by the Lexile Analyzer, referring to the total number of words in a text that is being analyzed. Not surprisingly, the results in Table 4 reveal that the reading texts in the test materials become longer as the educational stage moves up. Having said that, two observations warrant our attention here. First, there is a much more noticeable increase in the length of reading texts from senior secondary (M = 274.99, SD = 33.72) to collegiate basic (M = 506.10, SD = 288.70) than the gaps between other adjacent levels (e.g., from collegiate basic to collegiate intermediate). Second, the standard deviations for both collegiate basic (288.7) and collegiate intermediate (324.8) are substantial, suggesting significant variations in the length of reading texts used at these two educational stages. The large standard deviations of the word counts at the two collegiate levels are attributable to the nature of the tasks in the reading sections of CET4 and CET6, where both careful and expeditious reading tasks are used to assess students' reading ability. The reading texts in an expeditious reading task are much longer than those in a careful reading task. For the other two tests, however, expeditious reading is not included. It should be noted that word count does not influence the Lexile measures of the texts analyzed.

Complexity measure	Educational stage	n	Mean	SD	Q1	Q2	Q3	IQR
Mean Log Word	Junior secondary	200	3.66	0.17	3.54	3.69	3.79	0.25
Frequency	Senior secondary	200	3.55	0.16	3.45	3.54	3.64	0.19
	Collegiate basic	200	3.37	0.13	3.27	3.37	3.47	0.20
	Collegiate intermediate	200	3.30	0.12	3.22	3.30	3.38	0.16
Mean Sentence	Junior secondary	200	12.08	2.71	10.44	11.79	13.50	3.07
Length	Senior secondary	200	16.22	3.91	13.76	16.13	18.35	4.59
	Collegiate basic	200	19.19	3.43	16.68	19.30	21.04	4.36
	Collegiate intermediate	200	19.57	3.87	17.09	19.29	22.11	5.02
Word Count	Junior secondary	200	216.28	24.22	204.00	216.00	226.00	22.00
	Senior secondary	200	274.99	33.72	258.00	272.50	285.00	27.00
	Collegiate basic	200	506.10	288.70	342.00	348.50	492.50	150.50
	Collegiate intermediate	200	606.60	324.80	428.30	445.00	501.00	72.80
Lexile Measure [*]	Junior secondary	200	681.10	171.00	550.00	610.00	817.50	267.50
	Senior secondary	200	981.10	188.10	862.50	1000.00	1100.00	237.50
	Collegiate basic	200	1175.80	137.70	1072.50	1170.00	1277.50	205.00
	Collegiate intermediate	200	1215.00	137.80	1140.00	1210.00	1307.50	167.50

Table 4: Descriptive statistics of complexity measures generated by Lexile Analyzer across the four educational stages

Notes. SD: standard deviation; Q1: the 25th quartile; Q2: the 50th quartile; Q3: the 75th quartile; IQR: Inter-Quartile Range. ^{*}Lexile Measure should always have a capital L. For the sake of presentation, L is not included in this table.

The most important text complexity index generated by the Lexile Analyzer is the Lexile measure, representing the reading demand of a text in terms of the semantic difficulty and syntactic complexity. As mentioned, the Lexile measure ranges from below 0L (beginning readers) to above 1600L (advanced readers). The mean Lexile measures of the texts at the four educational stages are presented in Table 4. In addition, Figure 2 also graphically illustrates the distribution of text complexity across the four educational levels in focus, indicated by the median (the bar across each boxplot) and the IQR (Inter-Quartile Range, the upper and lower ends of each boxplot). The whisker of each boxplot represents 5th and 95th percentile. Figure 2 indicates that as the educational level progresses, the median of the Lexile measure also increases from junior secondary (610L) to senior secondary (1000L), next to collegiate basic (1170L) and finally to collegiate intermediate (1210L). Nonetheless, the increases from junior secondary and from senior secondary to collegiate basic tend to be larger than that from collegiate basic to collegiate intermediate. Another finding worth mentioning is that the Lexile measures vary significantly within each educational level, though junior and senior secondary levels tend to have larger variations than do the other two levels, that is, collegiate basic and collegiate intermediate.





To facilitate the comparison between the Lexile measures at each of the four educational stages in focus and the reading demands in other educational contexts, we present the median Lexile measures (with IQR) at each educational stage in Table 5 below, together with its corresponding level on the CSE. According to the CSE (Ministry of Education, 2018), Level 3 is aligned with the English proficiency of those who have completed junior secondary school, Level 4 senior secondary school, Level 5 the first two years of university study, and Level 6 the completion of the university education (Jin, et al., 2017).

Educational stage	CSE Level	Lexile measure			
Edubational stage		Median	IQR		
Junior secondary	3	610L	550L, 820L		
Senior secondary	4	1000L	860L, 1100L		
Collegiate basic	5	1170L	1070L, 1280L		
Collegiate intermediate	6	1210L	1140L, 1310L		

Table 5: Median lexile measures at the four educational stages

Note: According to the Lexile Analyzer User Manual, all Lexile measures are rounded to the nearest 10L to avoid overinterpretation of the measures.

5.1.2 Comparing the Lexile measures at the four educational stages

To compare whether a significant difference existed between the Lexile measures at the four educational stages, we first checked the distribution of the Lexile measures at each educational level. Figure 3 below displays the histogram of the Lexile measures at each level, indicating that they generally meet the normal distribution requirement. An examination of the skewness and kurtosis values for the data at each level indicates that they are all within the range of -2 to +2, thus supporting the normal distribution hypothesis (Field, 2013; Field, Miles, & Field, 2012). Though Levene's test indicates that the variances for the data at the four levels are not equal (p < 0.05), it was decided to proceed with one-way ANOVA because this statistical procedure is fairly robust to unequal group variances when sample sizes are equal, which is the case for this analysis (Field et al., 2012, p. 413).





The means of the Lexile Measures at the four levels, together with their 95% confidence interval, are displayed in Figure 4. ANOVA results indicate that the Lexile measures at the four levels differ significantly (F(3, 799) = 464.01, p < 0.05). A post hoc Tukey test shows that except for the insignificant difference between collegiate basic and collegiate intermediate (p > 0.05), the other differences between means in the pairwise comparisons were all significant (p < 0.05), suggesting that the Lexile measure at the collegiate basic level was larger than that of the senior secondary level, which was in turn larger than that of the junior secondary level.



Figure 4: Means of the Lexile measures at the four educational levels (with 95% CI)

5.1.3 Using Lexile measure to predict educational level

To explore the relationship between Lexile measures and educational level, we performed an ordinal logistic regression analysis, with the Lexile measure of a benchmark text as the predictor variable and the educational stage from which a text was selected as the dependent variable (DV). The ordinal logistic regression analysis was adopted because the DV in this analysis is an ordinal variable, with four levels (i.e. junior secondary, senior secondary, collegiate basic, and collegiate intermediate). Prior to the regression analysis, we calculated the correlation coefficient between the two variables. Spearman correlation was adopted in this case because one variable was continuous (the Lexile measure), whereas the other was ordinal (the educational level) (Field, 2013). The analysis indicated a strong relationship between the two variables, with a large effect size (Spearman's rho = 0.75, 95% CI: 0.71-0.78).

The ordinal logistic regression analysis results indicated that the predictor variable in this analysis (i.e., the Lexile measure) contributed significantly to the model ($\chi^2 = 797.78$, df = 110, p < 0.001), suggesting that the Lexile measure had a significant effect on determining the likelihood of a benchmark text being classified in one of the four educational stages. Pseudo R², which is interpreted in a similar way to R² in a linear regression analysis, was 0.67, suggesting that 67% of the variance in the DV could be explained by the predictor variable. The scatterplots in Figure 5 below illustrate the relationship between the predictor variable, that is, the Lexile measure, and the probability of a benchmark text being classified into one of the four educational levels.



Figure 5: Scatterplots of the Lexile measure and the probability of a benchmark text at each level

The scatterplot at the top left of Figure 5 illustrates that with the increase of the Lexile measure of a benchmark text (the *x* axis), the probability of it being classified in the junior secondary level decreases because it is more likely to be classified into a higher level. For the senior secondary level (the top right of Figure 5), however, it is clear that first, with the increase of the Lexile measure, the probability of a benchmark text being classified into this level also increases, until it reaches 930L (the peak in the bell-shape curve), after which the probability declines. A similar pattern is observed of the collegiate basic level (the bottom left), where the peak represents 1150L on the *x* axis.

Finally, for the collegiate intermediate level (the bottom right), the pattern is the opposite to that of the junior secondary level. With the increase of the Lexile measure, the probability of a text being classified into this level increases accordingly. These analysis results clearly demonstrate that the Lexile measure of a benchmark text could well predict the educational level from which a text is derived.

5.2 Comparing text complexity results of the Lexile Analyzer and Eng-Editor

5.2.1 Text complexity results generated by Eng-Editor

As mentioned previously (see Section 4.4), for each text that is being analyzed, Eng-Editor generates three text complexity values: lexical complexity, syntactic complexity, and text complexity, respectively. Each value consists of two parts: the complexity level of a text in accordance with the CSE levels before the decimal and the percentile of a text among all the texts at its assigned complexity level in the corpus in terms of its complexity. For example, if Eng-Editor produces a value of 3.68 for the lexical complexity of a text, it should be interpreted in light of two parts: 3 before the decimal represents that this text is at Level 3 on the CSE with regard to its lexical complexity and 68 after the decimal means this text is on the 68th percentile among all the texts at CSE Level 3 in the corpus that was built by the developer. In other words, this text is more difficult in terms of lexical complexity level assigned by Eng-Editor in terms of lexical, syntactic and text complexity is the most crucial result provided by this analysis tool. As such, we mapped the text complexity levels of the 800 benchmark texts assigned by Eng-Editor and the educational stages from which these texts were selected. Table 6 below summarises the mapping results.

As noted, the four educational stages of interest in this project correspond to Levels 3–6 on the CSE (i.e., junior secondary – CSE Level 3; senior secondary – CSE Level 4; collegiate basic – CSE Level 5; collegiate intermediate – CSE Level 6). Table 6 indicates that with regard to overall text complexity, there is a high level of alignment between the text complexity levels assigned by Eng-Editor and the educational levels, ranging from 88.0% (collegiate basic) to 97.0% (junior secondary). The finding suggests that Eng-Editor provides an accurate allocation of the overall text complexity level to each text, based on its algorithms. The mapping results regarding lexical complexity present a similar picture, with alignment percentages from 87.0 (collegiate basic) to 97.0 (junior secondary).

Having said that, the results in relation to syntactic complexity reveal a somewhat different picture, where the complexity levels of a quite high proportion of texts are not accurately aligned to the educational levels. This is particularly true for the benchmark texts at the two higher educational levels, with an alignment rate of 63.0 for collegiate basic and 77.0 for collegiate intermediate. It should be noted that for all misaligned cases, only two were misaligned by two levels, whereas the rest were by one level.

Table 6: Mapping the complexity levels assigned by Eng-Editor and educational levels

Complexity	Educational stage		n	Lovel alignment	Level misalignment (n/%)		
measure		CSE Level		(n/%)	Overestimate (n/%)	Underestimate (n/%)	
	Junior secondary	3	200	194 (97.0)	6 (3.0)	0 (0.0)	
Text	Senior secondary	4	200	186 (93.0)	4 (2.0)	10 (5.0)	
complexity	Collegiate basic	5	200	176 (88.0)	12 (6.0)	12 (6.0)	
	Collegiate intermediate	6	200	191 (95.5)	4 (2.0)	5 (2.5)	
	Junior secondary	3	200	194 (97.0)	6 (3.0)	0 (0.0)	
Lexical	Senior secondary	4	200	181 (90.5)	4 (2.0)	15(7.5)	
complexity	Collegiate basic	5	200	174 (87.0)	15 (7.5)	11 (5.5)	
	Collegiate intermediate	6	200	188 (94.0)	5 (2.5)	7 (3.5)	
	Junior secondary	3	200	197 (98.5)	3 (1.5)	0 (0.0)	
Syntactic	Senior secondary	4	200	161 (83.5)	2 (1.0)	37 (18.5)	
complexity	Collegiate basic	5	200	126 (63.0)	17 (8.5)	57(28.5)	
	Collegiate intermediate	6	200	154 (77.0))	5 (2.5)	41 (20.5)	

Notes: 'Level alignment' means the complexity level assigned by Eng-Editor matches the corresponding CSE level for an educational stage, whereas the reverse is true for 'level misalignment'. 'Overestimate' means the Eng-Editor level is higher than the target CSE level, and 'underestimate' means the opposite. In addition to mapping the results, we also correlated the complexity levels of the benchmark texts assigned by Eng-Editor and the educational levels from which they were selected. Given that both variables are ordinal in this analysis, Spearman correlation was adopted. Table 7 below presents the correlation results, with 95% confidence intervals. As indicated in this table, all variables are strongly and significantly correlated with each other. The correlation coefficients between the three complexity levels assigned by Eng-Editor and the educational level are all significant (p < 0.001), with large effect sizes (0.922-0.975).

Table 7: Correlation I	matrix between	Eng-Editor	complexity	levels a	ind educ	ational le	evels
(n = 800)							

	Text complexity	Lexical complexity	Syntactic complexity	Educational level
Text complexity	1.000			
Lexical complexity	0.986***	1.000		
	(0.983, 0.988)			
Syntactic complexity	0.921***	0.917***	1.000	
	(0.907, 0.932)	(0.902, 0.929)		
Educational level	0.975***	0.971***	0.922***	1.000
	(0.971, 0.979)	(0.966, 0.976)	(0.909, 0.934)	

Note: *** *p* < 0.001.

Figure 6 below illustrates the correlation between the Eng-Editor overall text complexity level and the educational level, with the regression fit line (dotted). As demonstrated in this figure, a few texts were misaligned at each of the four levels; however, the relationship between the two variables was strong. These analysis results generally supported the dependability of Eng-Editor as a criterion measure used to evaluate the accuracy of the analysis results of the Lexile Analyzer.

Figure 6: Correlations between educational level and text complexity level



5.2.2 Comparing text complexity results of the Lexile Analyzer and Eng-Editor

To compare the text complexity results generated by the two analysis tools, we first correlated the Lexile measures of the 800 benchmark texts used in this study to the complexity level assigned by Eng-Editor to each text. In this analysis, Spearman correlation was adopted because one of the variables, that is, the text complexity level is an ordinal variable. The result indicates that the two variables were significantly correlated (p < 0.001), with a large effect size (Spearman rho = 0.759, 95% CI: 0.723-0.790), suggesting a strong relationship between the overall text complexity measures generated by the two analysis tools.

Figure 7 below maps the Lexile measures of the benchmark texts onto their text complexity values generated by Eng-Editor across the four educational levels, with the regression fit line for each level. A linear regression analysis was implemented to investigate the extent to which the Lexile measures of a benchmark test could predict its complexity value that Eng-Editor produces. In this analysis, the predictor variable was the Lexile measure and the DV was the text complexity value provided by Eng-Editor. The results indicate that the model was significant (F(1, 799) = 133.49, p < 0.001), suggesting that the Lexile measure of a text contributed significantly to the text complexity values produced by Eng-Editor. Adjusted R² was 62.51%, suggesting that the Lexile measures can explain 62.51% of the variance in the DV (i.e. the Eng-Editor text complexity value).



Figure 7: Mapping Lexile measures with Eng-Editor text complexity levels

6. SUMMARY AND DISCUSSION

In this study, we investigated the extent to which the Lexile Framework was applicable to the Chinese educational context through analyzing 800 reading texts that have been used in four prominent, high-stakes English language tests at four educational stages in China, namely the English component of SHSEE, NMET, CET4 and CET6. To further interrogate the accuracy of the Lexile measures, we also compared the text complexity results generated by the Lexile Analyzer and Eng-Editor, a locally developed online text complexity analysis tool in China. Here, we discuss the major findings that emerged from our analyses in accordance with the two RQs that were investigated in this study.

RQ1 explored the range of Lexile measures for the benchmark texts at the four educational stages of focus, including junior secondary, senior secondary, collegiate basic, and collegiate intermediate. Our analysis with the Lexile Analyzer revealed that as expected, the Lexile measures increased incrementally as the educational stage progresses from junior secondary to collegiate intermediate (see Figure 2), signalling the increasing reading demands at a higher educational stage. The findings suggest that the Lexile Analyzer is generally effective in mapping the increasing complexity and reading demands at the four educational stages, aligned at Levels 3 to 6 on the China's Standards of English (CSE), the national English language proficiency framework in China. Nonetheless, it is worth noting that the reading demands, while increasing significantly from junior secondary to senior secondary and then from senior secondary to collegiate basic, seemed to level off from collegiate basic to collegiate intermediate (see Figure 2). This observation was confirmed through a subsequent analysis which compared the mean Lexile measures at the four levels, indicating that the increase of reading demands was statistically significant from junior secondary to senior secondary and from senior secondary to collegiate basic to collegiate intermediate (see Figure 2).

The two tests included in this study at the collegiate basic and collegiate intermediate level are CET4 and CET6. The development of both tests, including the selection of texts for their reading sections, follows rigorous quality control procedures (Jin & Yang, 2018). In addition, the validity of both tests has been supported by sustained validation research endeavours in the past three decades since the inception of the CET programme in the late 1980s (e.g., Fan & Frost, 2021; Jin & Yang, 2018). Therefore, we argue that this finding should not be attributed to the quality of the reading texts in the two tests; rather, it could be interpreted in light of the four educational stages that were included in this study. In the Chinese educational context, junior secondary school, senior secondary school and university represent three distinct educational stages, each marking a milestone in a student's educational trajectory (Jin et al., 2017); the two levels of collegiate basic and collegiate intermediate, however, are both at the university stage, with the former signifying basic and the latter higher requirements of CE learning in China's tertiary education (Jin & Yang, 2018).

Indeed, the benchmark texts at these two levels were selected from the same test (i.e. the CET), albeit at two levels (i.e., Band 4 and 6), whereas those at the other two lower levels (i.e. junior secondary and senior secondary) were selected from the English component of the SHSEE (*Zhongkao*) and NMET (*Gaokao*) respectively. Therefore, it is not entirely surprising that the increase of reading demands from collegiate basic to collegiate intermediate is not as significant as that from junior secondary to senior secondary and from senior secondary to collegiate basic. This is also implicated by the significant increase of the length of the benchmark texts at the two collegiate levels as compared with the two lower levels (see Table 2).

Another way of interpreting this finding is to refer to the descriptors of reading ability at Levels 3 to 6 on the CSE, which are the focus of this study. According to the CSE (Ministry of Education, 2018, p. 9), English learners at Level 3 (i.e., junior secondary) can locate key information in linguistically simple forms of writing and understand the implicit meaning and summarise the main points of short, linguistically simple materials on familiar topics. The descriptors at Level 4 (i.e., senior secondary), however, suggest a guite substantial increase in reading demands, as exemplified by the ability to understand the relationship between ideas by analyzing the structures of sentences and discourse whilst reading materials of medium linguistic difficulty. Similarly, at Level 5 (i.e. collegiate basic), there is a clear emphasis on reading materials characterised by medium linguistic difficulty containing opposing argumentation (e.g., editorials, book reviews), which is understandable as students are transitioning from senior secondary school to university study. The reading demands are therefore much higher as compared with the senior secondary level (i.e. Level 4). Nonetheless, the shift from Level 5 to Level 6 seems to be more characterised by the reading subskills required than the difficulty of the reading materials per se. For example, at Level 6, learners are expected to infer the writer's mood and attitude, whereas at Level 5, the key reading skills are differentiate facts and opinions and make simple inferences and understanding the relationship between ideas.

Given the wide application of the Lexile Framework in other contexts, it would be interesting to situate the reading demands at the four levels in focus in the global context (see Table 1 for a summary of the results in some contexts). It should be noted, however, most previous studies targeted English teaching materials, instead of English test papers. As revealed by our analysis, the median of the Lexile measure for junior secondary school is 610L (IQR: 550L-820L), largely consistent with Choi et al. (2012), who focused on the Lexile ranges of some English textbooks used in South Korea (510L–850L). As mentioned previously, we found only one study, MetaMetrics (2016), which analyzed the Lexile measures of the Eiken tests at different grade levels in Japan (see Table 2 for the analysis results). At the junior secondary level, the Lexile range of this study (median = 610L, IQR: 550L–820L) appears higher than that of MetaMetrics (2016, 530L–570L); however, at the senior secondary level, the results of the two studies seem to align reasonably well (this study: median = 980L, IQR: 860L–1100L; MetaMetrics (2016, IQR: 740L–1020L, see Section 2.1).

In addition, it is also meaningful to put our analysis results vis-à-vis the Lexile ranges of the textbooks in English, used in contexts where English is used as the first language (see Table 1). Not surprisingly, the textbooks in English in the US (850L–1110L) and UK (845L–1055L) are characterised by higher reading demands than the English test papers in China at the junior secondary level. However, it should be noted that at the senior secondary level, the reading demands appear to be quite comparable (US: 920L–1220L; UK: 920L–1110L; China: 860L–1100L). Furthermore, the reading demands at the collegiate intermediate level, as revealed by this study (IQR: 1140L–1310L), seem to align well with textbooks in English in the universities in the US (1220L–1390L) or UK (1210L–1360L).

RQ2 compares the text complexity results generated by the Lexile Analyzer, which implements the Lexile Framework, and Eng-Editor, a locally developed text complexity tool, with a view to further interrogating the accuracy of the Lexile measures. Generally speaking, the mapping results in Table 6 supported the dependability of Eng-Editor in classifying the benchmark texts into the four educational levels in focus (overall text complexity: 88.0%-97.0%). This conclusion was buttressed by the subsequent correlation analysis between the complexity levels assigned by Eng-Editor and the educational levels from which the benchmark texts were selected. However, it should be noted that quite a number of benchmark texts were misclassified at the two collegiate levels when it comes to syntactic complexity. This finding could be interpreted by the plateau effect of syntactic complexity for EFL learners, which is reported by Jin et al. (2020). They analyzed the syntactic complexity of over 3,000 adapted English texts used in 12 primary and secondary grade levels in China to map the progression of eight measures of syntactic complexity over time. They found that syntactic complexity measures tended to reach a plateau after EFL learners have attained a certain proficiency level. In addition, China's English instruction at the junior and senior secondary level has a clear emphasis on grammar, whereas CE teaching aims to train students' general and academic English abilities in the four essential skills (i.e., listening, reading, writing and speaking). As such, the syntactic complexity of the texts at the two collegiate levels may not be readily distinguishable from those at the lower levels. This might explain why Eng-Editor becomes less sensitive to the differences with regard to syntactic complexity at the two collegiate levels. Having said that, the correlation analysis between the Lexile measures and the overall text complexity levels assigned by Eng-Editor suggests a strong relationship between the text complexity results generated by the two analysis tools, hence supporting the accuracy of the Lexile Analyzer in evaluating text complexity in the Chinese educational context.

As mentioned previously, despite their focus on semantic difficulty and syntactic complexity, the theories and working mechanisms that underpin the Lexile Analyzer and Eng-Editor are different. The Lexile Analyzer was originally developed to evaluate reading development among L1 learners in the United States (Stenner, 1996), whereas Eng-Editor was developed by referring to local English curriculum standards, aiming to facilitate the selection and adaptation of English reading materials for local teaching and assessment purposes (Jin & Lu, 2018).

The strong relationship between the analysis results of the two tools lends empirical support to the applicability of the Lexile Analyzer to the Chinese educational context. In other words, an international framework is also able to identify progress in the local proficiency levels in the CSE. One might question the value of introducing the Lexile Analyzer to English language educators in China, given that Eng-Editor is an effective local tool for evaluating text complexity, as demonstrated by this study. However, we argue that the application of an internationally-recognised text complexity system, such as the Lexile Framework, is instrumental in situating the reading demands of different educational levels in China in the global context, which we believe is vitally important to the increasing internationalisation of English language education and assessment in China (e.g., Fan, Frost, & Jin, 2022). This is evidenced by, for example, the frequent academic exchanges between China and the rest of the world as well as a growing number of Chinese students studying in English-speaking countries. The Lexile Framework also functions as a link between the Chinese English language system and the rest of the world, providing further evidence for the wider relevance and use of the CSE and associated tests.

To date, the Lexile Framework, as mentioned previously, is not widely known in China, and the Lexile measure remains an alien notion to most English language educators, learners and assessment developers in China. The findings of this study lend empirical support to the applicability of the Lexile Framework to the Chinese educational context, at least in terms of the reading texts in high-stakes English tests.

In view of the findings of this study, we encourage English language teachers, materials writers and assessment developers in China to actively engage and apply the Lexile Framework in their professional practices. We also encourage language assessment researchers to leverage the strengths of the Lexile Framework in their research. One fruitful direction, for example, is to analyze and compare the Lexile measures of the English textbooks at different educational levels, which can then be compared with those of the test materials, thus enhancing the quality of English teaching materials as well as the validity of English reading assessments in China.

REFERENCES

Bush, L., Koons, H., & Sanford-Moore, E. E. (2018a). *Comparing the reading demand of the English CSAT and English textbooks in the Republic of Korea*. Metametrics Research Brief. Retrieved from: <u>https://metametricsinc.com/wp-content/uploads/2018/01/MM English CSAT and English</u> <u>Textbooks in the Republic of Korea Digital.pdf.</u>

Bush, L., Koons, H., & Sanford-Moore, E. E. (2018b). *Primary and secondary English textbook complexity in the Republic of Korea*. Metametrics Research Brief. Retrieved from: https://metametricsinc.com/wp-content/uploads/2018/01/MM Textbook complexity in Korea Digital.pdf

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing*, 25(1), 15–37.

Choi, S., Kim, K. H., Lee, Y. B., Hong, J. H., & Cho, E. (2012). Developing the Korean EFL Readability Formula (KRF). *English Language & Literature Teaching*, *18*(3), 1–24.

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge: Cambridge University Press.

Dong, M., Fan, J., & Xu, J. (2021). Differential washback effects of a high-stakes test on students' English learning process: evidence from a large-scale stratified survey in China. *Asia Pacific Journal of Education*, 1–18.

Fan, J. (2018). A survey of English language testing practices in China: Students' and teachers' perspectives. In D. Xerri & P. V. Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 283–300). Cham, Switzerland: Springer.

Fan, J. (2019). Assessing Chinese learners of English: language constructs, consequences and conundrums. *Assessment in Education: Principles, Policy & Practice, 26*(3), 356–361.

Fan, J., & Frost, K. (2021). At the intersection of language testing policy, practice, and research: An interview with Yan Jin. *Language Assessment Quarterly*, 1–14.

Fan, J., Frost, K., & Jin, Y. (2022, online first). Local English testing at the tertiary level in China: Contexts, policies, and practices. *Language Testing.* DOI: 10.1177/02655322211070839

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Thousand Oaks: Sage.

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using R. Thousand Oaks: Sage.

Guo, K., Chen, J., Lei, J., & Jin, T. (2021). Teacher engagement with technology-enhanced text adaptation for reading assessment: A case study. *International Journal of Computer-Assisted Language Learning and Teaching, 11* (4), 100–112.

Jin, T., Guo, K., Mak, B., & Wu, Q. (2017). Lexical profiles of reading texts in high-stakes tests: Where are the benchmarks? *International Journal of Computer-Assisted Language Learning and Teaching*, *7*(1), 34–49.

Jin, T., Li, Y., & Li, B. (2016). Vocabulary coverage of reading tests: Gaps between teaching and testing. *TESOL Quarterly*, *50*(4), 955–964.

Jin, T., Li, S., & Liu, F. (2021). Selection and adaptation of academic reading texts. *Foreign Language World, 202* (1), 20–27.

Jin, T., & Lu, X. (2018). A data-driven approach to text adaptation in teaching material preparation: Design, implementation and teacher professional development. *TESOL Quarterly*, *52*(2), 457–467.

Jin, T., Lu, X., & Ni, J. (2020). Syntactic complexity in adapted teaching materials: differences among grade levels and implications for benchmarking. *The Modern Language Journal, 104*(1), 192–208.

Jin, Y., & Yang, H. (2018). Thirty years of the College English Test (CET). *Foreign Language World, 2*: 29–39.

Jin, Y., Wu, Z., Alderson, C., & Song, W. (2017). Developing the China Standards of English: challenges at macropolitical and micropolitical levels. *Language Testing in Asia, 7*(1).

Koons, H., Bush, L., & Sanford-Moore, E. E. (2018). *Primary and secondary textbook complexity in England.* Metametrics Research Brief. Retrieved from: <u>https://metametricsinc.com/wp-content/uploads/2018/01/MM Primary Secondary England Research Digital.pdf</u>

Lee, B. (2014). 당신의 영어는 왜 실패하는가? [Why does your English learning fail?] Seoul, Korean: 우리학교.

Lennon, C., & Burdick, H. (2004). The lexile framework as an approach for reading measurement and success. Retrieved from: <u>https://cdn.lexile.com/m/resources/materials/Lennon__Burdick_2004.pdf</u>.

MetaMetrics. (2016). Lexile measurement of tests: EIKEN & Test of English for Academic Purposes. Retrieved from: <u>https://metametricsinc.com/research-publications/lexile-measurement-tests-eiken-test-english-academic-purposes/</u>

Minitab, LLC. (2021). Minitab. Retrieved from: <u>https://www.minitab.com</u>.

Ministry of Education. (2018). China's Standards of English Language Ability. Retrieved from: http://www.neea.edu.cn/res/Home/1908/0c96023675649ac8775ff3422f91a91d.pdf.

Negishi, M. (2015). Lexile Measure による中高大の英語教科書のテキスト難易度の研究[Lexile Measures of Textbooks Used in Secondary and Tertiary Education in Japan]. *ARCLE Review*, 9, 6-16. <u>https://www.arcle.ip/research/books/data/html/data/pdf/vol9_3-1.pdf.</u>

Spolsky, B. (1995). Measured words. Oxford: Oxford University Press.

Stenner, A. J. (1996). Measuring Reading Comprehension with the Lexile Framework.

Supplemental information for Appendix A of the Common Core State Standards for English language arts and literacy: New research on text complexity. (n.d.). Retrieved from: <u>www.corestandards.org/wp-content/uploads/Appendix-A-New-Research-on-Text-Complexity-revised.pdf.</u>

Van Moere, A., & Spiby, R. (2017). *Understanding the reading demands in an English medium instruction program.* Paper presented at the 5th New Directions English Conference. Shanghai: China.

Wang, L. (2008). The spread of English in China and its implications. *Australian Review of Applied Linguistics*, *31*(3), 321–324.

Wilkins, C., Hartman, J., Howland, N., & Sharma, N. (2010). How prepared are students for collegelevel reading? Applying a Lexile®-based approach. *Issues & Answers Report, REL*(094). Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2013). The Common Core State Standards' quantitative text complexity trajectory: Figuring out how much complexity is enough. *Educational researcher*, *4*2(2), 59–69.

Williamson, G. L., Koons, H., Sandvik, T., & Sanford-Moore. (2012). *The text complexity continuum in grades 1–12*. Metametrics Research Brief. Retrieved from: <u>https://metametricsinc.com/wp-content/uploads/2017/09/Research Brief-Text Continuum Grades 1-12 1.pdf</u>.

Williamson, G. L., Sandvik, T., Stenner, J., & Johnson, A. (2015). *Complexity of university texts in the United Kingdom*. Metametrics Research Brief. Retrieved from: <u>http://cdn.lexile.com/cms_page_media/147/UK%20Text%20Research%20Brief_2015_8.5x11.pdf#:~:t</u> <u>ext=The%20median%20text%20complexity%20of%20texts%20used%20in,United%20Kingdom%20a</u> <u>nd%20the%20United%20States%20of%20America</u>.

Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing, 38*(2), 219–246.

Zheng, Y., & Cheng, L. (2008). Test review: College English Test (CET) in China. *Language Testing*, 25(3), 408–417.



Research into Reading Grants

APPLYING THE LEXILE FRAMEWORK TO EVALUATE TEXT COMPLEXITY IN THE CHINESE CONTEXT: A CORPUS-BASED STUDY

RESEARCH IN EFL READING AND READING ASSESSMENT Report 1, 2022

Jason Fan, Tan Jin & Ivy Chen

BRITISH COUNCIL www.britishcouncil.org

METAMETRICS The Lexile Framework for Reading – Lexile Matching Readers with Texts www.lexile.com

© British Council and MetaMetrics 2022 ISSN 2755-2543