

# Matching EFL learners with appropriate levels of reading materials: Backing for using Extensive Reading Placement/Progress Test

EFL 学習者の言語能力レベルに適した多読マテリアル：  
EPER テストの妥当性研究

Kiyomi Yoshizawa  
吉澤清美

英語多読では語彙・構文が学習者の言語能力に適した読み物を数多く読む。これら多読用読み物はグレーデッドリーダーと呼ばれる。学習者の英語能力とその能力に適した読み物を一致させることは多読授業を成功させる重要な要因の一つである。本研究では、英語を第2言語として学ぶ学習者の読みのレベルを測定するために開発された Edinburgh Project on Extensive Reading Placement/Progress Test (EPER\_ppt) が英語を外国語として学ぶ学習者の読みのレベルを測定するテストとして使うことの妥当性を Bachman & Palmer (2010) 並びに Bachman & Damböck (2017) で提唱されている枠組みを使い、検証し、その証拠を提示した。

## キーワード

backing, Edinburgh Project on Extensive Reading Placement/Progress Test, extensive reading, graded readers, validity

## 1. Introduction

### 1.1 The aim of the study

Extensive reading (ER) has been widely recognized as an effective way to learn languages for the last two decades. However, matching learners with reading materials suitable for their reading ability is a crucial factor for a successful ER program. Previous research indicates that if less than 90% of words on one page of the reading material are known to a learner, the reading material becomes too difficult to comprehend. This is called the “frustration level.” In

this case, the learner must repeatedly look up unknown words in a dictionary. This results in overwhelming the learner's cognitive capacity and reading with less comprehension and slower fluency. This kind of reading is more likely to demotivate the learner to read. If the percentage of unknown words on one page of the reading material is between 90 to 98%, it is called the "instructional level." The learner reading at the instructional level can make inferences on the context around unknown words and understand the major points of the reading material. If a precise understanding is required, they still need to use a dictionary for checking the meanings of unknown words. When the rate of unknown words on one page of the reading material is 98% or more, the material is appropriate for the learner's reading ability and he/she can read it with comprehension and fluency (The Extensive Reading Foundation, 2011). It is called the "independent level." The current study aims to present evidence for using the Edinburgh Project on Extensive Reading Placement/Progress Test (henceforth, EPER\_ppt) to match EFL (i.e., English as a foreign language) learners at a university in Japan with reading materials suitable for their reading ability.

## **1.2 How teachers make sure that their students read extensive reading materials appropriate for their language abilities**

Some institutions use standardized English proficiency tests such as TOEFL or TOEIC (e.g., Nishizawa, Yoshioka, & Nagaoka, 2017; Nishizawa, et al., 2011) when they determine appropriate levels of their students' reading materials. However, standardized test results are not readily available in certain situations. Other ER instructors ask their students to decide the appropriate levels of their graded readers by examining the number of unknown words on a page: If a student finds only "a small number of unknown words (no more than two or three words per page) and they understand almost everything, then that is a suitable book (Nation & Waring, 2020, p. 58)." Similar suggestions were made by Waring (1997) and Wan-a-rom (2010). Although self-assessment is widely accepted among the ER practitioners, there are some drawbacks. One is students' guessing of the meaning of an unknown word may depend on the frequently used meaning in a bilingual or monolingual dictionary, but it may not match the intended meaning, resulting in misunderstanding the text. Further, without any means of assessing learners' progress, it is difficult to observe such progress learners make in reading with comprehension and fluency over a period of instruction.

Other ER teachers use EPER\_ppt for placing learners and determining their progress. (Institute for Applied Language Studies University of Edinburgh, 1990, 1995). EPER\_ppt has been developed by the Edinburgh Project on Extensive Reading to match English language

learners with their appropriate reading materials.

### 1.3 What is EPER\_ppt?

EPER\_ppt is a cloze test developed to measure general English proficiency, including grammar in context, and to place learners into appropriate reading levels. The cloze test, by its nature, requires learners to process text at various levels, ranging from clauses to sentences. EPER\_ppt comprises 12 short passages (approximately 80 words on average) taken from different levels of obsolete graded readers and arranged in an ascending order of difficulty. There are 141 deleted items at the rate of 4 to 12 words. As the 12 short passages in the EPER\_ppt were all taken from obsolete graded readers, the test is considered as an appropriate test to measure general English proficiency and the improvement of the learners who read various graded readers for ER.

### 1.4 EPER readability levels

EPER developed a database of high-quality language learner literature. EPER rated English graded readers in terms of “clarity of the plot, the support of illustrations, the appearance of printed page, and most important, interest of the story or book” (Day & Bamford, 1998, p. 172). The graded readers in the EPER database are classified into nine levels of readability. These levels are developed by EPER “to provide a common scale on which to place books published from various series and various publishers” (p. 172). This is due to the fact that each publisher uses their own criteria of readability, resulting in a situation where levels of one publisher’s graded readers are not comparable to those of others. Thus, this common scale called the EPER scale provides a common yardstick for assessing readability of language learners’ reading materials. Further, the EPER scale helps “the learners to make easy progress from level to level” (p. 172).

The EPER readability scale has nine levels (i.e., EPER levels). Day and Bamford (1998) summarized the EPER levels and the corresponding levels within the major English language proficiency tests such as IELTS and TOEFL. The students’ ages are grouped into three groups: a) primary students, including those aged 8 to 11, b) secondary students, including those aged 11 to 16, and c) adults, including those over the age of 16. They further list about 550 titles available in the world market which EPER graded high according to their assessment criteria. Table 1 presents a simplified version of the information presented in Day and Bamford with the recommended graded readers (Institute for Applied Language Studies, University of Edinburgh, 2002). The table includes the EPER levels, average vocabulary, and the series and

publishers of recommended graded readers. In the table, the abbreviations of the books are as follows: CER, *Cambridge English Readers* (secondary to adult); MMR, *Macmillan Readers*; MMRL, longer versions of *Macmillan Readers*; OBW, *Oxford Bookworms Library* (secondary to adult); OBWS, *Oxford Bookworms Starters* (secondary to adult); PGR, *Penguin Graded Readers* (secondary to adult); PGRE, *Penguin Graded Readers Easy Starts* (lower than level 1 of PGR); PYR, *Penguin Young Readers* (primary). The number following an abbreviation of a book refers to the levels of the series which each publisher uses.

Table 1 *EPER Levels, average vocabulary, and recommended graded readers*

Levels	Average vocabulary	Recommended Graded Readers
H	*	PYR1, PYR2
G	300	PYR3, PGRE, OBWS, MMR1
F	500	PGR1, PYR3, MMR2
E	800	MMR2L, CER1, OBW1
D	1,200	PGR2, OBW2, MMR3, CER2
C	1,600	PGR3, OBW3, MMR4, CER 3
B	1,900	PGR4, OBW4, MMR4, MMR5, CER4
A	2,200	PGR5, OBW5, CER5
X	3,000	PGR6, OBW6, CER6

Note. \* information is not provided.

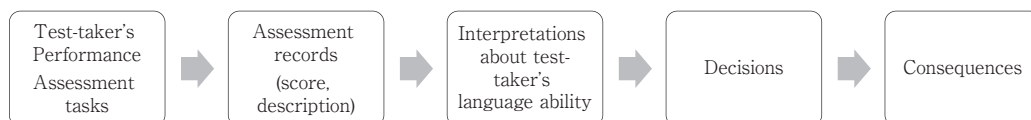
One of the issues raised by instructors using EPER\_ppt is that the EPER levels resulting from EPER\_ppt are higher than the reading levels of English language learners in English as a foreign language context (EFL) such as in Japan (Takase, personal communication, November 16, 2021). As mentioned above, the EPER scale was developed for English language learners in Europe, an English as a second language (ESL) context. It needs to be confirmed that the EPER scale is appropriate for learners in an EFL context.

The present study aims to examine to what extent EPER\_ppt is appropriate as a placement test for EFL learners at the university level in Japan. The current study utilizes the framework for language assessment development and uses by Bachman and Palmer (2010) and Bachman and Damböck (2017). We present Claim 3, a statement that the interpretation of the ability assessed by EPER\_ppt is relevant to the formative decision (i.e., placement of learners to their appropriate reading levels with comprehension and fluency) and backing (i.e., evidence) to support our claim.

### 1.5 Validation of the EPER\_ppt

According to Bachman and Palmer (2010) and Bachman and Damböck (2017), a language assessment involves four inferential links from the test takers' performance on assessment tasks to the consequences the assessment has on its stakeholders. Figure 1 shows the inferential links from assessment performance to consequences.

Figure 1 *Inferential links from assessment performance to consequences*



Note. Adapted from L. Bachman and A. Palmer, 2010, p. 91.

The first inferential link lies between the test-takers' performance on the assessment task and the assessment records. The second inferential link lies between the assessment records and the interpretation of the test-takers' language ability. Based on the information on the assessment records, decision makers interpret the performance on the assessment task as an indicator of the test-takers' language ability. The third link lies between the interpretation of the test-takers' language ability and the decisions made by decision-makers. The interpretation of the test-takers' language ability provides information to make decisions about them. The last inferential link lies between decisions and consequences. The decisions that the decision-makers make have consequences to stakeholders. In cases high-stakes decisions are involved, the consequences may bring life-affecting influences to stakeholders such as admission to higher educational institutions or failure in a course. In cases of low- to mid-stakes decisions, learners may modify the focus of their learning or teachers may make changes in their teaching methods or materials.

Bachman and Palmer (2010) state that developers and users of the assessment for decision-making need to consider the stakeholders who will be affected by the decisions made based on such assessments (p. 85). They further state that "in order to be accountable, test developers and decision makers must demonstrate through argumentation and the collection of supporting evidence, that the use of a particular assessment is justified" (pp. 85-86). Bachman and Palmer define accountability as the ability to demonstrate to stakeholders that "the intended uses of our assessments are justified" (p. 86).

An assessment use argument (AUA) is defined as "a series of claims or statements" which describe the links from a test-taker's performance on an assessment to the intended consequences of using the assessment (Bachman & Damböck, 2017, p. 30). Based on Bachman and

Palmer (2010) and Bachman and Damböck (2017), the author created Claim 3 for EPER\_ppt, which states:

The interpretation of the students' reading ability of narrative text with appropriate understanding and fluency is relevant for the formative decision to be made (i.e., placement of learners to their appropriate reading levels with comprehension and fluency). The interpretation is sufficient for the low-stakes decisions to be made. The interpretation is meaningful for a reading class where the participants are expected to develop reading ability with general comprehension and fluency. The interpretation is generalizable to their current language program and impartial to all students.

The next step of the assessment use justification was to collect backing. Backing is defined as "evidence that the test developer and test user need to provide in order to support the claims in AUA" (Bachman and Damböck, 2017, p. 36). Bachman and Damböck further explain that there are two kinds of backings. The first kind concerns procedures followed while an assessment is developed or used. The second kind of backing is "collected during the assessment development and use, specifically for the purpose of supporting claims in the AUA" (p. 36), including "analyses of students' performance, using a range of qualitative and quantitative approaches, as appropriate to the particular claim in the AUA" (p. 36). According to Bachman and Damböck (2017), backings for Claim 3 include reports on the results of additional assessments or assessment tasks.

The aim of the current study is to present the backing (i.e., evidence) to support Claim 3 (interpretation). The current study reports a classroom-based language assessment focusing on EPER\_ppt as a placement test for the learners in a reading program for initiating an extensive reading portion of the program. Extensive reading accounts for at least 25% of the final grades of the learners in the program. The syllabus of the reading program aims to develop learners' reading comprehension with fluency. Consequently, the backing of Claim 3 in the current study focuses on comprehension and fluency. The backing of comprehension consists of two sources. The first is the number of unknown words in the reading text, which is recommended based on the EPER\_ppt scores. The second is how the learners at each level perceive the readability level of the recommended books. Furthermore, the backing of fluency also consists of two sources. The first source is the reading rates at which each learner is capable to read their recommended readers. The second source is the extent to which each learner translates their recommended reader into their L1. Each of these sources of backing is

explained in detail in the Methods section of the current paper. The rest of the paper presents a method for collecting backing for Claim 3 (interpretation), results, and discussion, followed by educational implications and limitations of the study.

## 2. Methods

### 2.1 Backing for Claim 3 (interpretation): reading graded readers at the EPER-recommended level and lower levels

To collect backings for Claim 3, three 10-minute reading tasks were created. In the first week of the first semester, participants took EPER\_ppt. In the following week, they were asked to read the graded reader, which was recommended by the EPER scale (henceforth, RL graded reader), for ten minutes based on their performance on the test. After the 10-minute reading, they were asked to respond to five follow-up tasks which were designed to elicit information about the following five points: the reading amount during the 10-minute reading, unknown words in their reading, the readability level of their readers, the amount of L1 translation during the 10-minute reading, and their comments about the readability of their readers. After the five follow-up tasks were completed, the 10-minute reading task and follow-up tasks were repeated for the graded reader one level lower than the EPER recommended level (henceforth, 1L graded reader) and for the graded reader two levels lower than the recommended level (henceforth, 2L graded reader).

The first of the follow-up tasks was to record the reading progress during the 10-minute reading. The total number of words read in the book was indicated by writing the page number, line number, and the word in the section of the book they read when the 10-minute call was announced. Based on the information provided by each participant, we calculated the reading amounts of the readers the participants read and reading rates in words per minute (henceforth, WPM). The reading rates for graded readers at EPER-recommended levels were compared and contrasted with those of L1 readers who read grade-appropriate texts in terms of readability. We use L1 readers' reading rates as one criterion against which we can judge the fluency level of EFL readers in the current study.

The second task was to list unknown words and the page numbers where they appeared for the first time. This provides the information about the characteristics of unknown words in each graded reader. To analyze the words in each graded reader, lexical profiles were created for each reader. All the graded readers were scanned and scanned pages were converted into text files using AntFileConverter (Anthony, 2017). All the prefaces, captions with illustrations,

comprehension and vocabulary questions after main stories were excluded from the analyses. Lexical profiles were created using Range Program (Heatley et al., 2002), which uses base-words of the first three 1,000 word families in British National Corpus.

Third, the readability level was measured by the participants, rating the readability level of each graded reader on a scale of 1 to 5 where 1 is *too easy* for the readers; 2 is *rather easy*; 3 is *appropriate* for their reading level; 4 is *a little difficult*; 5 is *difficult*. It was assumed that most of the participants would choose 3, if the readability levels of their graded readers fit their reading proficiency levels. In addition, we assumed that the participants would choose 2 (rather easy) or 1 (too easy) if the readability levels of the graded readers were lower than their reading proficiency levels. It was also assumed the participants would choose 4 (a little difficult) or 5 (difficult) if their graded reader was higher than their reading proficiency levels.

The fourth task asked the participants about the amount of translation during the 10-minute reading. We assumed that the participants would utilize their L1 as their resource when they read an L2 text which exceeded their appropriate reading levels. Kern (1994) examined the roles of mental translation during L2 reading and provided evidence that L2 learners use L1 translation as they “work through the meaning of difficult passages” (Grabe, 2009, p. 215). L1 translation helps especially weaker learners to have better understanding of the text (Kern, 1994). The participants in the current study had just completed six years of English instruction in the Japanese school system where L1 translation is one of their reading strategies. We assumed that the participants would use L1 translation more when they read RL graded readers compared to 1L or 2L readers. The translation question was on a five-point Likert scale and they were asked to select 1 (i.e., no L1 translation is made) to 5 (i.e., L1 translation is made all through a text).

The fifth task asked the participants to write what made a graded reader difficult or easy; for example, vocabulary, grammar, syntax, the relationships among characters in the book, and the background of the story described by a graded reader.

## 2.2 Participants

The participants of the current study were 19 first-year students at a university located in the western part of Japan. They were in a one-year reading class where most of the instruction was based on a designated reading textbook targeting English language learners at B1 and B2 of the Common European Frame of Reference. In their first year, they had 15 hours of English language instructions per week to develop their proficiency.

The class time was spent mostly to cover the content of the designated textbook. In each

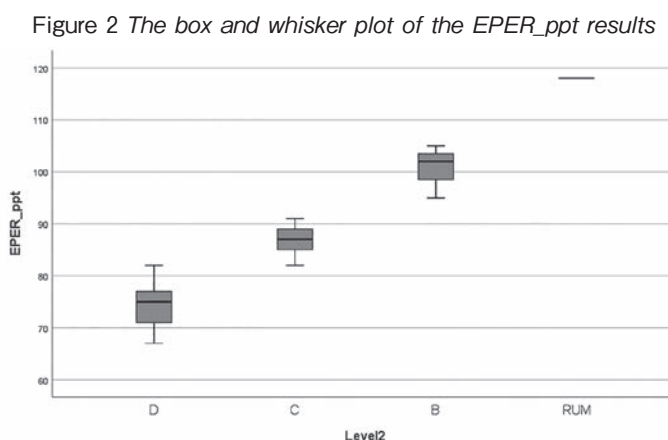


class, ten to fifteen minutes were spent reading extensive reading materials. In addition, the students performed most of their extensive readings outside the regular class hours.

## 2.3 Instrument

### 2.3.1 EPER\_ppt

In the first week of the first semester, the participants took EPER\_ppt. Based on the test results, students were assigned to levels D, C, B, and RUM. RUM stands for Reading Unlimited Materials. According to the EPER scale, the participant at this level can read, ungraded reading materials. Figure 2 shows the medians of EPER\_ppt of four levels.



### 2.3.2 Graded readers suitable for EPER levels D, C, B, and RUM

Based on the results of EPER\_ppt, three graded readers were selected for each level: a graded reader at the level recommended by the EPER scale (RL graded reader), a second reader, one level lower than the recommended level (1L graded reader), and a third reader, two levels lower than the recommended level (2L graded reader). For example, if a student's recommended level is D, graded readers from levels E and F were selected. For all the graded readers used in the study were from the list of the books recommended as those most suitable for secondary students aged 11 years or older. Table 2 shows the titles of the graded readers used in the current study, their series and levels, and readability indices. Two more readers were used for one participant at RUM level; however, they were not analyzed in the current study.

Two types of readability indices were used: YL and Flesch-Kincaid Grade Level (FKGL). YL stands for *yomiyasusa level*, subjective readability levels established by instructors who are practicing extensive reading in Japan. It has 10 levels and each level has 10 subdivisions. In the

current study, YL increased as the EPER levels increased and the differences in YL ranged from 0.6 to 0.7 between the two adjacent levels. FKGL is based on word and sentence lengths. It describes text readability as a grade in the United States (US) school system. It shows that the readability of the five readers increases as the EPER scale increases. Additional information about the five readers are included. SPP stands for sentences per paragraph. WPS stands for words per sentence and CPW stands for characters per word. WPS indicates that the number of words per sentence increases as the EPER scale levels increase.

Table 2 Information on the graded readers used in the study

Levels	Titles	Series	YL	FKGL	SPP	WPS	CPW
F	<i>The Long Tunnel</i>	MMR2	1.4	2.0	2.1	6.7	4.2
E	<i>Love or Money?</i>	OBW1	2.1	2.3	3.5	6.6	4.0
D	<i>Henry VIII</i>	OBW2	2.7	3.3	2.7	9.0	4.0
C	<i>Frankenstein</i>	OBW3	3.3	4.1	4.1	12.3	3.8
B	<i>The Unquiet Grave</i>	OBW4	3.9	5.1	3.0	14.8	4.0

Note. Only the first 4,756 words were analyzed in *the Unquiet Grave*.

For each reader, its Flesch-Kincaid Grade Level (FKGL) is used as an indicator of a corresponding grade level and the approximate reading rate is obtained from existing research on L1 reading fluency.

## 2.4 Data collection procedure

In the first week of the first semester, EPER\_ppt was administered in class for 45 minutes. The test papers were scored using the answer key. Based on the test performance, the reading levels of the participants were decided. The RL, 1L, and 2L graded readers and a follow-up task sheet were placed in an envelope for each participant. In the second week, the 10-minute-reading tasks were administered.

## 3. Results/Discussion

### 3.1 The number of words read in ten minutes

Table 3 presents the number of words the participants read in the graded readers at three readability levels. Considering level D participants, the medians of words increased as the participants read the graded readers from RL to 1L and 2L. On the other hand, the median of level C participants decreased when they read 1L graded reader, *Henry VIII*, but they read the

2L reader, *Love or Money?*, more than the RL graded reader, *Frankenstein*. Concerning level B participants, their reading amount increased when they read 1L reader, *Frankenstein*; however, the reading amount decreased when they read 2L reader, *Henry VIII*. These results indicate that some of the characteristics of *Henry VIII* slowed the reading of the participants. However, we can still see the tendency that the number of words read increases as the readability levels of the graded readers become lower. On the other hand, within each level, the ranges of the number of words the participants read are wide. For example, the minimum number of words read at the RL for level D participants is 932 and the maximum number of words by their counterpart is 1609.

Table 3 *The number of words read in ten-minute readings*

	Two levels lower	One level lower	Recommended level
Level D	<i>The Long Tunnel</i>	<i>Love or Money?</i>	<i>Henry VIII</i>
<i>Mean</i>	1590.2	1357.5	1269.8
<i>Median</i>	1437.0	1306.5	1279.0
<i>Min.</i>	962.0	885.0	932.0
<i>Max.</i>	2523.0	1972.0	1609.0
Level C	<i>Love or Money?</i>	<i>Henry VIII</i>	<i>Frankenstein</i>
<i>Mean</i>	1582.1	1532.9	1448.8
<i>Median</i>	1537.0	1415.0	1483.0
<i>Min.</i>	928.0	995.0	1122.0
<i>Max.</i>	2551.0	2265.0	1840.0
Level B	<i>Henry VIII</i>	<i>Frankenstein</i>	<i>The Unquiet Grave</i>
<i>Mean</i>	1632.0	1660.0	1635.3
<i>Median</i>	1504.0	1675.0	1552.0
<i>Min.</i>	1420.0	1304.0	1527.0
<i>Max.</i>	1972.0	2001.0	1827.0

### 3.2 The reading rates of the participants

Table 3 in the previous section shows the means and medians of the number of words the participants read in 10-minute readings. One tenth of the total number of words read was considered to be the participants' WPMs. The median of Level D participants was 127.9 for reading the RL graded reader. Similarly, the medians of Level C and B participants were 148.3 and 155.2 WPMs, respectively.

Several previous L1 reading studies have presented reading rates of native speakers of English at different grade levels. For example, Rasinski (2003) lists words correct per minute (wcpm) for oral reading rates by grade. Carver (1992) lists standardized words per minute for

silent reading by grade. In the current study, we used words per minute (WPM) for silent reading reported by Spichtig et al. (2016). They used WPM to assess the comprehension-based silent reading efficiency of the elementary and junior high school students in the United States. Their study indicated that second graders read 115 WPM; fourth graders, 147 WPM; sixth graders, 165 WPM; eighth graders, 166 WPM; tenth graders, 185 WPM; and twelfth graders, 192 WPM. We used the midpoint between the second and fourth grades' reading rates, 131 WPM, as an estimate for the reading rate for the third graders and the midpoint of the fourth and sixth graders' reading rates, 156 WPM, as an estimate of the reading rate of the fifth graders.

In the current study, *Henry VIII*, the graded reader recommended for the level D participants, has the Flesch-Kincaid Grade Level 3.3 and the median reading rates of the level D participants is 127.9, which is equivalent to the reading rate of L1 readers reading grade-appropriate text. On the other hand, the data show that the variability among the participants' reading rates was large: the range was 67.7 words. Similarly, the level C participants read the text at a rate comparable to the US fourth graders, 147 WPM. The median of the level C participants reading the RL graded reader, *Frankenstein*, was 148.3 WPM. However, the range was 72.8 words and the variability among the level C participants was large. Even though there were only three level B participants, a similar tendency was observed in them. The median reading rate was 155.2 WPM, which is very close to the estimate of reading rates of the US fifth graders, 156 WPM. We can conclude that the participants in the current study read at the rates appropriate for each readability level, though the reading rates varied among the participants at each level.

However, reading rates do not necessarily reflect the readability levels of text. We assume that the participants could read text with lower readabilities faster than the recommended graded readers. Figures 3-5 show the WPMs of the participants at levels D, C, and B, respectively, when they read the 2L, 1L, and RL graded readers. Y axis indicates reading rates. X axis indicates the readability (YL) of each graded reader. Thick straight lines with black round markers in the middle of each figure show the mean WPMs. In some cases (i.e., D1, D3, D4, C2, C4, and C9), the patterns of their reading rates meet our assumption. In the other cases (i.e., D2, D5, D6, C1, C3, C5, C7, C8, and B3), the reading rates of the 1L graded readers are lower than those for reading the RL books. Similarly, C5-C8 and B1-2 read the 2L graded readers slower than the 1L graded readers. Further, some readers read graded readers at almost the same rate, regardless of their readability levels. For example, D2 read books at a rate less than 100 WPM irrespective of whether he/she read the graded readers appropriate for his/her

language proficiency. These results suggest that reading rates are influenced by factors unrelated to linguistic characteristics of graded readers.

Figure 3 Level D learners' reading rates

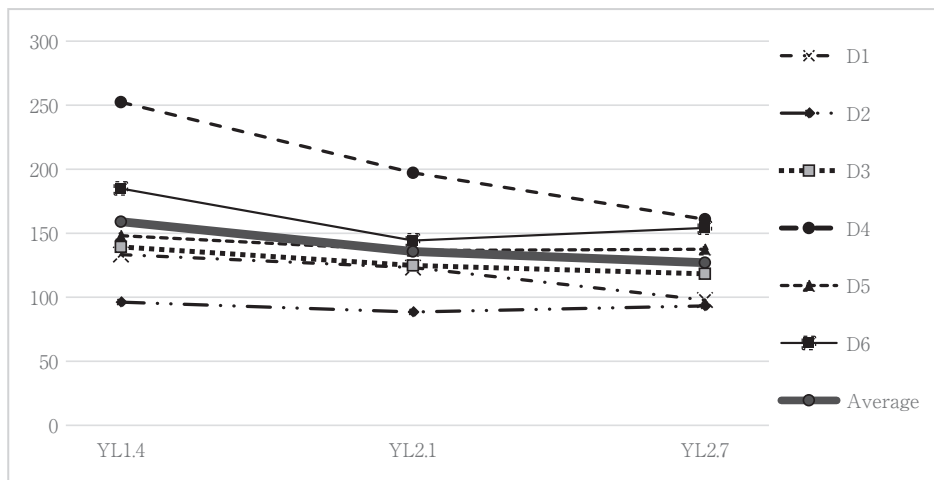


Figure 4 Level C learners' reading rates

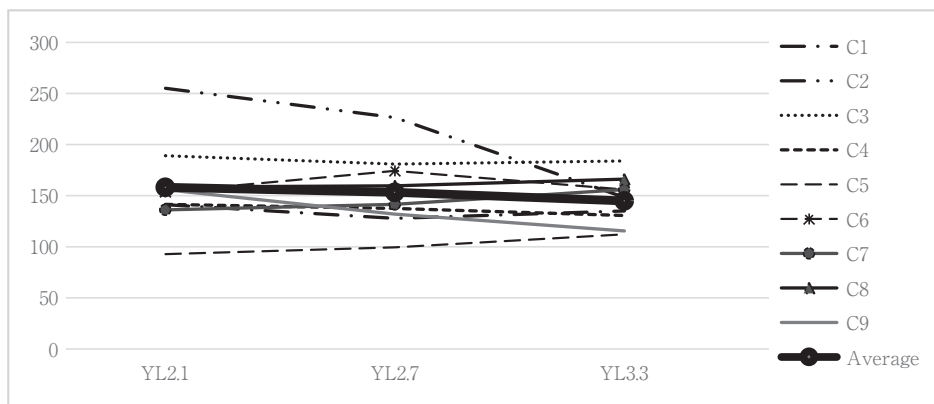
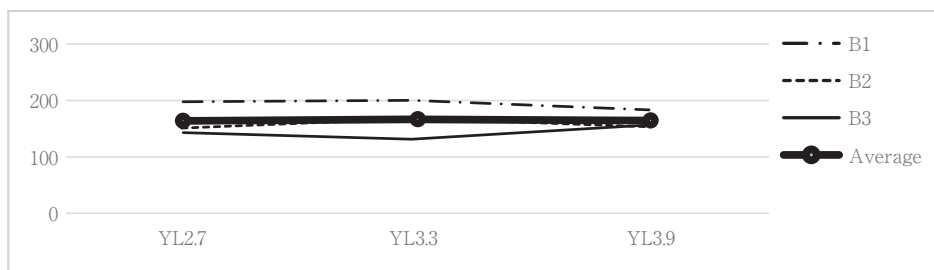


Figure 5 Level B learners' reading rates



### 3.3 The percentage of unknown words

The unknown words the participants listed while reading each reader were analyzed. Table 4 lists the words the participants marked as unknown words. In the second column from the right, the first numbers in parentheses indicate the number of the participants who wrote the listed words. In the following combinations of letter B and the number, B stands for baseword and the following number indicates the frequency bands, that is, whether a word is listed in the first, second, or third most frequent 1,000-word families. NF stands for not found, indicating the words are not listed in the first to third most frequent 1,000-word families. For example, in *the Long Tunnel*, “banknotes” is reported by two participants and it is not included as the most frequent 3,000 words. “Branch” and “Lorry” are reported by one participant, respectively, and included in the second most frequent word list. The last column shows the percentage of unknown words, which is calculated by dividing the number of tokens of these unknown words by the total number of words in each graded reader. The results indicate that the percentage of unknown words is less than 1% and it provides further backing that the readability levels of these graded readers were appropriate for levels D and C participants. Since no word was listed as unknown by the level B participants, we assume that the recommended reading level was also appropriate for them.

Table 4 *The list of unknown words*

Levels	Titles	Unknow words	%
F	<i>The Long Tunnel</i>	banknotes (2, NF), branch (1, B2), lorry (1, B2)	0.30%
E	<i>Love or Money?</i>		—
D	<i>Henry VIII</i>	miscarriage (1, NF), beheaded (9, NF), mistress (2, NF), riding (1, B2), palace (3, NF), Windsor (1, NF), Pope (1, NF)	0.57%
C	<i>Frankenstein</i>	mast (4, NF), sledge (1, NF), faint (2, B3), learnt (1, B1)	0.37%
B	<i>The Unquiet Grave</i>		—

### 3.4 How the learners perceive the readability levels of their ER materials

Table 5 shows the frequency and percentage of the self-assessment of the readability levels of the RL, 1L, and 2L graded readers. The numbers 1-4 below “Self-assessment of readability levels” are responses the participants marked. For the RL graded readers, 68.4% of the participants assessed that they matched with their reading proficiency levels and 31.6% assessed that they were rather easy for them. For the 1L graded readers, 57.9% of the participants assessed that they matched with their reading proficiency levels and 36.8% assessed that they were

rather easy for them. However, one participant (5.3%) assessed that the readability of the 1L graded reader was rather difficult. The participant read *Love or Money?* and her written comment indicates that the difficulty is not related to a language problem, but the complexity of the relationships among the characters in the story. This point is further explained in the last section.

Concerning the 2L graded readers, 15.8% of the participants assessed the readers at this level as too easy for them and 36.8% assessed them as rather easy. 31.6% assessed that the readers at this level matched their reading proficiency levels. On the other hand, 15.8% thought the graded readers were rather difficult. These responses came from the level C participants who read *Love or Money?* as their 2L graded reader. In their comments, the same source of difficulty mentioned above was listed.

Although the number of the participants was small, a general tendency was observed in the self-assessment results about the readability levels of the graded readers they read. There is no clear distinction in the patterns of the levels D and C participants' assessments between the RL and 1L readers. Furthermore, when the participants read the 2L readers, the number of the participants who considered the readability of the graded readers would match their reading

Table 5 *The self-assessment of the readability levels of the RL, 1L, and 2L graded readers*

Participants	Self-assessment of readability levels			
	1	2	3	4
Graded readers at recommended levels				
D	—	3 (50.0)	3 (50.0)	—
C	—	2 (22.2)	7 (77.8)	—
B	—	1 (33.3)	2 (66.7)	—
RUM	—	—	1 (100.0)	—
Total	—	6 (31.6)	13 (68.4)	—
Graded readers at one level lower				
D	—	2 (33.3)	3 (50.0)	1 (16.7)
C	—	2 (22.2)	7 (77.8)	—
B	—	3 (100.0)	—	—
RUM	—	—	1 (100.0)	—
Total	—	7 (36.8)	11 (57.9)	1 (5.3)
Graded readers at two levels lower				
D	1 (16.7)	3 (50.0)	2 (33.3)	—
C	1 (11.1)	1 (11.1)	4 (44.4)	3 (33.3)
B	1 (33.3)	2 (66.7)	—	—
RUM	—	1 (100.0)	—	—
Total	3 (15.8)	7 (36.8)	6 (31.6)	3 (15.8)

proficiency level (i.e., the number of participants who chose 3 on the scale) decreased and the participants considered the 2L graded readers to be rather easy or too easy for their reading ability (i.e., the number of the participants who selected 2 or 1 on the scale).

### 3.5 The amount of translation

For the level D participants, 50% of them translated the RL reader fairly often, but this tendency decreased as they read the 1L and 2L readers. For the level C participants, 67% of them made little or no L1 translation when they read the RL and 1L readers. When they read the 2L reader, about 78% of the level C participants made little or no L1 translation. We can see that less translation is utilized when the readability levels of the graded readers increase. On the other hand, the amount of translation of level B participants did not vary, regardless of text readability. Two level B participants made little or no L1 translation, but one used L1 translation quite often or all the time.

### 3.6 Participants' comments

The participants' comments provide hints about factors that make graded readers easier or more difficult for them to read. In general, the participants commented that the RL graded readers were easy to read. Although there were few unknown words for the participants, they did not pose any problems in understanding the RL readers. Furthermore, grammar or sentence structures were not problematic for them. In Excerpt 1, D4 comments that she did not find the grammar and vocabulary in her RL reader as difficult and the reader was easy to read. Similar comments were made across all three levels, resulting from the EPER\_ppt.

Excerpt 1 (D4)

Bunpoya tangoga sorehodo muzukashikunaku, yomiyasukatta

(Grammar and vocabulary are not so difficult and the recommended graded reader is easy to read.)

Excerpt 2 (C1)

Shiranai tango, bunpoga hotonodo nakatta

(There were few unknown words or difficult grammatical points.)

Excerpt 3 (B1)

Tangomo bunpomo chodo yokatta



Matching EFL learners with appropriate levels of reading materials (Yoshizawa)

(Vocabulary and grammar are appropriate.)

The complexity of the relationships among the characters in the graded readers posed some difficulty to the participants. For example, *Henry VIII* is the RL book for the level D participants, the 1L reader for the level C participants, and the 2L reader for the level B participants. It is a narrative told by the sixth wife of Henry VIII of England, Catherine Parr, and her maid, Margaret. They read letters written by the first five wives. As the participants read this particular reader, they had to figure out the relationships among the characters and this posed them some difficulty. Excerpts 4 and 5 demonstrate this point. Similar comments were made across all three levels.

Excerpt 4 (D4)

Henrino tsumaya koga takusan detekurunode sukoshi yayakoshikatta

(The story describes many characters who are Henry's wives and children, which complicates this story a little complicated.)

Excerpt 5 (B1)

tango ha yasashi keredo tojyojinbutsuga ookute yayakoshikatta

(Although vocabulary is easy, many characters appear, which makes the story complicated.)

Similarly, *Frankenstein* poses some difficulties due to the complexity of the relationships among characters.

### 3.7 Summary of the major findings with regard to backing for using EPER\_ppt

In terms of relevance, the extent to which “the intended interpretations provide the information the test user needs to make the decision” (Bachman & Damböck, 2017, p. 67), the results of the present study are likely to present backing for using EPER\_ppt as an assessment tool for matching learners with appropriate extensive reading materials. As explained in Section 1.5, backing or evidence to support the use of EPER\_ppt is based on two aspects: comprehension and fluency. In terms of comprehension, we examined unknown words in the EPER-recommended graded readers and the participants' self-assessment of the readability levels of graded readers at recommended and lower levels. The percentage of unknown words in the graded readers at the recommended and lower levels is less than 1%. This confirms that the

vocabulary levels of each graded reader meet the requirement for 98% coverage as readers at the independent level. Further, this result is reflected in the participants' assessment of graded readers they read in the 10-minute reading tasks. For the participants at all levels, the recommended graded readers were suitable or easy for them. Similar response patterns were observed for reading the graded readers one level lower across all three levels. Concerning the graded readers two levels lower than the recommended levels (2L graded readers), the majority of the participants at levels D and B perceived their readers as easy or too easy. On the other hand, the non-linguistic factor in the 2L graded readers made those less readable for the level C participants.

On the one hand, fluency results indicate that the participants read the RL readers at rates equivalent to those of third, fourth, and fifth graders. This finding demonstrated that RL graded readers were suitable for the reading abilities of the participants. On the other hand, 47.4% of the participants made L1 translation "sometimes" or "often." This result provides an inconclusive basis for considering the recommended readers as suitable for each participant. This was probably because some participants might have depended on L1 translation for comprehending L2 text.

Finally, the comments written by the participants indicated that the linguistic characteristics of the graded readers were not the only cause of difficulties in reading graded readers in the 10-minute reading tasks.

The decisions to be made based on the results of EPER\_ppt were formative, that is, using the information to help students choose graded readers which match their reading abilities and for teachers to use the information to assist learners in choosing the graded readers suitable for their reading levels and advise them on what materials are available. Considering the results of the present study, the EPER\_ppt provides a sufficient source of information which the participants and their instructor need regarding the suitability of a graded reader in relation to the students' language proficiency when they start an extensive reading program.

### **3.8 Educational implications**

The results of the self-assessments of readability of graded readers provide educational implications. One is that learners' self-assessment of the readability levels of graded readers could differ from their actual reading levels and the difference could sometimes be one grade apart. In many ER programs, learners choose the books. Some learners use their interests as the criteria for selecting books. Sometimes, learners may choose graded readers appropriate for their reading abilities; in other times, graded readers are far more difficult to read. It may be

best for an instructor to inform learners of the range of books from which they can select and monitor their progress. Additionally, the current study shows that the reading rates of EPER levels D, C, and B are equivalent to those of the third to fifth graders in the US school context. If some of the learners read much slower than those reading rates, they are more likely to have some problems in reading. They might be translating every L2 text sentence by sentence, regardless of its readability level. The instructor can use approximate reading rates as a sign of students' reading problems. In this sense, it is important for students to make logs of some kind and for their teacher to check them on a regular basis.

### 3.9 Limitations of the study

The current study was not without limitations. First, the sample size was small. Second, the participants were in an intact group and generalizability was limited. EFL learners in different school contexts may result in different findings. Third, the comprehension of reading graded readers used in the 10-minute reading was not examined. This needs to be incorporated into further studies. Fourth, the self-assessment measure for checking readability levels of graded readers is based on one item and its reliability as a measurement is very limited. Likewise, the amount of translation is based on one item and its reliability is limited.

The current study has these limitations; thus further studies are needed with different groups of EFL learners and with a larger sample size.

### Acknowledgement

This study was supported by JSPS Grant-in-Aid for Scientific Research (C) Grant Number 17K02996. The author expresses a deep gratitude to Dr. Atsuko Takase for all the support given to the completion of the current study.

### References

- Anthony, L. (2017). AntFileConverter (Version 1.2.1) [Computer Software]. Waseda University. Available from <https://www.laurenceanthony.net/software>
- Bachman, R., & Damböck, B. (2017). *Language assessment for classroom teachers*. Oxford University Press.
- Bachman, R., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Carver, R. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, 36, 84-95.
- Day, R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge

- University Press.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Heatley, A., Nation, I.S.P. and Coxhead, A. (2002). RANGE and FREQUENCY programs. [Computer Software]. Victoria University of Wellington. Available from <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>
- Institute for Applied Language Studies University of Edinburgh. (1990). *E.P.E.R. Edinburgh Project on Extensive Reading Placement Test Testpack A (Complete)*. University of Edinburgh.
- Institute for Applied Language Studies University of Edinburgh. (1995). *E.P.E.R. Edinburgh Project on Extensive Reading Placement Test Testpack A (Complete)*. University of Edinburgh.
- Institute for Applied Language Studies University of Edinburgh. (2002). *E.P.E.R. Edinburgh Project on Extensive Reading List of Graded Readers Current Titles March 2003*. University of Edinburgh.
- Kern, R. (1994). The role of mental translation in L2 reading. *Studies in Second Language Acquisition* 16, 441-61.
- Nation, I. S. P., & Waring, R. (2020). *Teaching extensive reading in another language*. Routledge.
- Nishizawa, H., Yoshioka, T., & Nagaoka, M. (2017). How many words should elementary EFL learners read extensively and from which readability levels. *Toyota Kogyo Senmon Gakko Kiyo* [Journal of Toyota National Institute of Technology], 50, 1-12.
- Nishizawa, H., Yoshioka, T., Ito, K., Nagaoka, M., Hiroyama, S., & Asai, H. (2011). Critical factors in a successful extensive reading program of English. *Journal of Japanese Society for Engineering Education*, 59(4), 66-71.
- Rasinski, T. (2003). *The fluent reader: Oral reading strategies for building word recognition, fluency, and comprehension*. Scholastic Books.
- Spichtig, A., Hiebert, E., Vorstius, C., Pascoe, J., Pearson, P., & Radach, R. (2016). The decline of comprehension-based silent reading efficiency in the United States: A comparison of current data with performance in 1960. *Reading Research Quarterly*, 51(2), 239-259. <https://doi.org/10.1002/rrq.137>
- The Extensive Reading Foundation. (2011). *Kokudai tadoku kyouikugakkainiyoru tadokushido gaido* [Extensive Reading instruction Guide compiled by the Extensive Reading Foundation].
- Wan-a-rom, U. (2010). Self-assessment of word knowledge with graded readers: A preliminary study. *Reading in a Foreign Language*, 22(2), 323-338.
- Waring, R. (1997). Graded and extensive reading—Questions and answers. Special issue: Extensive reading. *The Language teacher*, 21(5). Retrieved from [https://jalt-publications.org/tlt/issues/1997-05\\_21.5](https://jalt-publications.org/tlt/issues/1997-05_21.5)