# RESEARCH

# Free and Open Access

# Explainable eBook recommendation for extensive reading in K-12 EFL learning

Kensuke Takii <sup>1</sup>\*, Brendan Flanagan <sup>2</sup>, Huiyong Li <sup>3</sup>, Yuanyuan Yang <sup>4</sup>, Kento Koike <sup>5</sup> and Hiroaki Ogata <sup>6</sup>

\*Correspondence: kensuke.takii96@gmail.com Graduate School of Informatics, Kyoto University, Kyoto, Japan Full list of author information is available at the end of the article

# Abstract

An automatic recommendation system for learning materials in e-learning addresses the challenge of selecting appropriate materials amid information overload and varying self-directed learning (SDL) skills. Such systems can enhance learning by providing personalized recommendations. In Extensive Reading (ER) for English as a Foreign Language (EFL), recommending materials is crucial due to the paradox that learners with low SDL skills struggle to select suitable ER resources, despite ER's potential to improve SDL. Additionally, determining the difficulty level of ER materials and assessing learners' progress remains challenging. The system must also explain its recommendations to foster motivation and trust. This study proposes a mechanism to estimate the difficulty of ER materials, adapted to learner preferences, using information retrieval techniques, and an explainable recommendation system for English materials. An experiment was conducted with 240 Japanese junior high school students in an ER program to assess the accuracy of difficulty estimation and identify learner characteristics receptive to the recommendations. While the recommendations did not significantly impact learners' English skills or motivation, they were positively received. A strong relationship was found between the use and acceptance of recommendations and learners' motivation. The study suggests that although the system did not increase overall motivation, it has potential to further enhance the motivation of naturally motivated learners.

**Keywords:** E-learning, English as a Foreign Language, Learning material recommendation, Extensive reading, System transparency

# Introduction

In the context of foreign language learning, Extensive Reading (ER) is a learning method in which students read many texts spontaneously to acquire information or for pure enjoyment (Day et al., 1998). In ER, it is essential for learners to choose materials that suit them and to read at their paces, and it is recognized that these are effective in helping



© The Author(s). 2024 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

learners acquire language. Indeed, ER effectively promotes foreign language skill development, especially in improving reading comprehension and vocabulary growth (Schmitt, 2008; Urquhart & Weir, 2014). The characteristic of ER, in which students are free to read books of their choosing at their own pace, is also closely related to learners' Self-directed Learning (SDL) strategies. For example, previous studies have shown that ER can influence the process of learners becoming self-directed learners (Enisa et al., 2013; Ningsih, 2019; Takahashi & Umino, 2020). Therefore, ER, like SDL, is closely related to attributes such as motivation and autonomy of individual learners.

However, the paradox lies in the difficulty of learners with low SDL skills to select appropriate ER books, even though ER is critical to enhancing these skills. This can be a problem, especially with online learning. Regarding e-learning, the number of new digital learning repositories continues to grow, and learners point to excessive online learning resources (Ochoa & Duval, 2009). In e-learning, therefore, it is vital to have an inclusive learning environment that allows various learners to select the materials appropriately.

Such difficulties in selecting appropriate materials/exercises commonly appear in any domain. To overcome this problem in general, several studies have been conducted on recommendation systems to automatically recommend the most appropriate learning materials for individual learners. However, several problems to be solved can be considered in the context of eBook recommendations in ER. First, it is challenging to make a good estimate of the learner's ability and recommend the right material. ER cannot apply recommendation algorithms, including estimation of the learner's ability and item's difficulty based on learners' response histories, such as Bayesian Knowledge Tracing (BKT) (Corbett & Anderson, 1994) and Item Response Theory (IRT) (Hambleton & Swaminathan, 2013). This is because it cannot define the "correct answer" in only reading activities unlike in general quizzes, and it is difficult to define what and to what extent students have learned with reading materials. It is also difficult to unambiguously determine the difficulty level of the material in learning English. The difficulty level of reading materials is determined by a combination of factors, including the English vocabulary comprising the materials (Kasim & Raisha, 2017; Qarqez & Ab Rashid, 2017), English grammar items, and the difficulty of the content itself (Satriani, 2018). Therefore, in the ER context, it is difficult to make recommendation algorithms with both estimations of the learner's ability and the item's difficulty, like existing research. Second, it is difficult to realize appropriate recommendations that are acceptable to learners. Recent research revealed that if the recommendations are black-boxed, the learner who does not trust will disagree (Abdi et al., 2020).

Due to the two above problems in the context of eBook recommendations in ER, this study addresses two key challenges. Firstly, we propose an algorithm for estimating the difficulty level of a book by aggregating the difficulty levels of the words within it. Secondly, assuming that learners' preferences reflect the difficulty levels of books previously read by learners, we realize appropriate recommendations in which the learner's characteristics match the difficulty of eBook. By adopting these two solutions, we propose an explainable English learning material recommendation system for ER in EFL. It can estimate learners' preferences from their ER activities and recommend materials that match their estimated preferences. In addition, we also present a mechanism for estimating the difficulty level of learning materials for ER for the realization of this system. The recommendation platform includes an eBook reader system and a vocabulary profile. Learners' usage logs of eBooks for learning materials and vocabulary difficulty levels are processed using TF-IDF, an information retrieval technology, to estimate the difficulty level of the materials and the learner's preference for the difficulty level of materials. Using the estimated difficulty level and preference, the system recommends learning materials of the appropriate difficulty level. The system can respond to chronological changes in the learner's preference by processing the learner's learning logs in real-time. In addition, the system can generate explanations that provide the basis for the recommended materials based on the proximity of their difficulty level to the learner's preference. This explanation can provide additional information to make the recommendation more persuasive.

This recommendation is intended to be a persuasive recommendation that adapts to the learner's preference and encourages spontaneous learning by the learner. However, as indicated above, this requires an accurate estimation of the difficulty level of the material. In addition, for this ER material recommendation to be accepted by more learners and to support learning effectively, it is necessary to understand the characteristics of the learners who supported it. Therefore, we pose the following two research questions corresponding to the above two defined challenges:

**RQ1:** To what extent does the proposed mechanism for estimating the difficulty level of English materials for ER estimate the difficulty level of the materials?

**RQ2:** What characteristics of learners supported the recommendation of English language materials for ER that could be explained?

# **Related work**

#### **Extensive reading**

Extensive Reading (ER) is a foreign language learning method in which learners read as many books as possible without focusing on unfamiliar words or phrases. Tanaka and Stapleton (2007) conducted a semi-extensive reading program at a Japanese junior high school where EFL reading is inadequate and found that ER significantly increased exposure to reading input. They concluded that ER had positive effects on English language learners. Previous studies on ER have included print (Mason & Krashen, 1997;

Mermelstein, 2015; Pitts et al., 1989; Tanaka & Stapleton, 2007) and online materials (Chen et al., 2013; Sun, 2003). Although integrating online materials into ER is less representative than traditional print text media (Chen et al., 2013), ER programs have proven to have similar positive effects on learners regardless of the type of medium. ER has also been shown to improve reading speed (Tanaka & Stapleton, 2007) and reading comprehension (Chen et al., 2013; Mason & Krashen, 1997; Tanaka & Stapleton, 2007), as well as writing (Mason & Krashen, 1997; Mermelstein, 2015) and vocabulary (Pitts et al., 1989).

ER is closely related to learners' SDL strategies because of the methodological feature of allowing learners to read freely and at their own pace. For example, ER has been shown to influence how learners become self-directed (Enisa et al., 2013; Ningsih, 2019; Takahashi & Umino, 2020). However, while ER is necessary to improve SDL skills, selecting appropriate ER materials for learners with low SDL skills is difficult. This is particularly problematic in the context of e-learning.

#### Personalized recommender systems in e-learning

In contrast to the general recommendation system, which assumes the existence of an accurate answer to a recommendation (Bahrani et al., 2024), personalized recommendations in e-learning systems require knowledge of the learner and the learning material (Shishehchi et al., 2011). While most state-of-the-art recommender systems in education/learning support uses learners' needs, learning styles, or preferences (e.g., Chen et al., 2024), such recommendation should consider the learning outcomes resulting from the recommendation (Sikka et al., 2012). The learner and material knowledge need to be modeled so that computers can handle the recommendation that can bring about appropriate learning effects. The task of modeling knowledge about learners can be rephrased as modeling learner features needed for personalized recommendations. Learner preferences (Bourkoukou & El Bachari, 2018; Hsu, 2008) and learning styles (Klašnja-Milićević et al., 2011; Truong, 2016) are examples of characteristics to be modeled. For example, an attempt was made to complement the recommender system with existing open learner models (Abdi et al., 2020).

Modeling the knowledge to be learned is another crucial step for personalized recommendations. The attribute-based material recommendation system developed by Salehi and Kmalabadi (2012) uses matrix preferences that reflect the attributes of the material and the learner's access log. Vocabulary recommendation by Zou and Xie (2018) also models lexical knowledge by using lexical sets to represent the difficulty of each word. Several studies use ontologies to model domain knowledge relevant to recommendations (George & Lal, 2019; Tarus et al., 2018). An ontology is an explicit knowledge representation format that shows knowledge items and their relationships in a domain;

Flanagan et al. (2019) developed a lexical knowledge map, reflecting the words in coursework materials and the semantic context in which they are used. This knowledge model has been used in EFL vocabulary recommendation for refugees (Abou-Khalil et al., 2021) and in an English picture book recommendation system (Takii et al., 2021).

However, it is difficult to automatically infer the learner's state of knowledge because of the difficulty in determining the difficulty level of materials and defining knowledge acquired through specific learning behaviors. Zou and Xie (2018) developed an explicit learner profiling model for personalized vocabulary recommendation, where learners' vocabulary size and proficiency level were modeled. It allows learners to adjust the recommended vocabulary level but not automatically infer it. Therefore, this research aims to provide recommendations tailored to learners' preferences regarding the difficulty of the material rather than estimating learners' knowledge states.

The EFL material recommendation in this study uses a content-based method that uses information retrieval techniques to automatically infer the difficulty level of the material and the learner's preference for it. This allows the recommendation system to simultaneously model the learner's knowledge and material knowledge. This method is also helpful for explanation generation, which is discussed in the following subsection.

#### Generation of explanations for recommendations in e-learning

Previous research has shown that intelligent tutoring systems with prompting and feedback mechanisms can increase students' motivation in self-regulated learning and lead to higher achievement (Duffy & Azevedo, 2015). Such systems with transparent mechanisms can be persuasive and trustworthy to learners, motivating them and leading to effective learning. A study by Ooge et al. (2022) on explaining exercise recommendations showed the need for transparency and trust in the recommendation system. Flanagan et al. (2021) proposed a system called EXAIT (Educational eXplainable Artificial Intelligent Tools), a system of e-learning systems that can help students learn effectively—proposed to address the issue of learner trust and motivation behind recommendations made by e-learning systems. Their recent work developed an educational mathematics exercise recommendation system using the Bayesian Knowledge Tracing algorithm (Takami et al., 2023) and knowledge concepts extracted from textbooks (Dai et al., 2022).

In ER, closely related to self-directed learning, the system must motivate the learner to perform better. Therefore, material recommendation systems used in ER should provide additional information on their recommendation and be transparent, explainable, and trusted by learners.

# Positioning of this study

Together with the context of EFL learning, this research aims to provide ER material recommendations tailored to learners' preferences regarding the difficulty of the material rather than estimating learners' knowledge states. This is because, as mentioned earlier, estimating learners' knowledge states from their learning behaviors is a difficult task. The EFL material recommendation in this study uses a content-based method, which uses information retrieval techniques to automatically infer the difficulty level of the material and the learner's preference for it. This allows the recommendation system to simultaneously model the learner's state and material knowledge. In addition, the recommendation system also aims to ensure transparency within the system by explaining the reasons for recommending the material. It is hoped that this feature of the recommendation system will enable learners to trust the system and consequently increase their motivation to learn.

## **Recommendation platform**

# **Recommender overview**

Figure 1 shows an overview of the use of the recommendation platform. First, the platform administrator prepares a vocabulary list and teaching materials from the teaching materials store in advance. This word list contains vocabulary difficulty information and is used to evaluate the difficulty level of the teaching materials. Next, learners use the educational materials on the eBook reader system. This system converts the reading logs of the materials into the format of the xAPI (Advanced Distributed Learning Initiative, 2013), a comprehensive repository of learning/educational records. The reading logs stored in the



LRS, along with the difficulty level information for each material, should be designed to be sent to the Learning Record Store (LRS) (xAPI.com, 2011), a comprehensive repository of learning/instructional records. Next, the recommendation mechanism generates the materials to be recommended to the learner, along with explanations of the reasons for the recommendations. These explanations include the weight of the recommendation (how much the material should be recommended) and the reasons for the recommendation. Finally, the generated recommendations are presented to the learner through the recommendation interface.

#### **Platform components**

Students can do the ER activity using the BookRoll eBook reader (Figure 2). BookRoll is designed to access eBooks or lecture slides inside or outside the classroom (Flanagan & Ogata, 2018). The BookRoll system tracked students' online operations, such as flipping to the next or previous pages. All reader operations are recorded in the LRS. With the help of BookRoll, students' reading behaviors are collected, including the reading pages, words, time, and speed.

The user interface for the recommendation shown in Figure 3 is implemented on the Goal-Oriented Active Learning (GOAL) system (Yang et al., 2024). The GOAL system is a platform to support students' development of data-informed SDL ability (Li et al., 2021; Majumdar et al., 2018). It provides five recommended learning materials at most. Users can jump directly to the BookRoll by clicking the title and can read the recommended material. Each recommended material is followed by the recommendation weight and explanatory sentences explaining why the recommendation was made to the learner. The provided explanation can be shown or hidden when learners click a button next to the explanation.





# **Recommendation mechanism**

#### Mechanism overview

The recommendation mechanism we propose in this paper uses model-driven and datadriven approaches. First, as a model-driven approach, this mechanism uses a wordlist with information on the difficulty of vocabulary, which is used to evaluate the difficulty of materials in the book library. Next, as a data-driven approach, reading logs estimate learners' preferences for the difficulty levels of materials with the estimated material difficulty. Then, the system makes material recommendations based on the material difficulty and the learner's preference. This mechanism generates recommendations so that the difficulty of recommended materials will be as close to the learner's preference as possible. Besides, the descriptions which explain why the material has been recommended to the learner are displayed.

# Material difficulty evaluation

#### Difficulty of vocabulary

The difficulty level of instructional materials is determined by several factors that make up the materials. In this study, we focus on vocabulary difficulty and assume that vocabulary primarily determines the difficulty of instructional materials. In other words, we assume that teaching materials containing complex vocabulary are generally tricky, and teaching materials containing a few problematic or many easy vocabulary words are easy to read. However, if the complex vocabulary is not essential for understanding the material, learners can understand the content without knowing the meaning. Therefore, to estimate the difficulty level of the material, we used the vocabulary profile and the ranking function in information retrieval.

This study used the CEFR-J Wordlist Version 1.6 (Tono, 2020) as a reference for vocabulary difficulty. This vocabulary list is based on the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001). It is a corpus created from English textbooks used in China, Taiwan, and Korea, from which common vocabulary used in CEFR-level texts in each country/region is extracted. It was constructed for English language education. The list contains 6868 headwords, each of which has four levels of difficulty: A1, A2, B1, and B2 of the CEFR.

To calculate the difficulty level of the material, we first quantified the difficulty level of each word as A1 to 1, A2 to 2, B1 to 3, B2 to 4, and so on. Since words with multiple meanings have multiple levels for each meaning, these difficulty levels were set as multiple averages. For example, "will" has difficulty level A1 as a modal auxiliary verb and level B2 as a noun meaning mental power. The difficulty is 2.5 in this case, with an average of 1 (A1) and 4 (B2). The exact words are also considered different words if they are spelled differently (e.g., "color" and "colour," both with difficulty A1). Therefore, the difficulty of the word *t*, represented as D(t), should be expressed as  $D(t) \in [1,4]$ .

# Difficulty of materials

The difficulty of the documents was calculated using each word introduced above, the TF-IDF score (Anand & Jeffrey, 2011), and the number of words comprising each document. In our method, all words in each document are extracted and the document is regarded as a bag of words. TF-IDF is a ranking function defined for a tuple of documents and words in information retrieval, which can be applied to bags of words (sets of words in which the elements are allowed to be duplicated) and estimates the relevance of a document to a query. The TF-IDF score refers to the relevance of a query word to a document and indicates the importance of the word in the document. In this study, all words in the "CEFR-J Wordlist Version 1.6" are assumed to be query words. The document *d* and the word *t* is high, *t* is an essential word for understanding the contents of *d*. Besides, we deal with each material as a bag of words consisting of the words in the vocabulary list. Namely, when a set of all words in the vocabulary list is represented as  $\Sigma$ , the material *d* is expressed as  $d' \cap \Sigma$ , where *d'* is a bag of words of the material corresponding to *d*.

The difficulty of a material d is calculated by using the following formula:

$$D(d) = f\left(c_1 \sum_{t \in d} D(t) \cdot \text{TFIDF}(d, t) + c_2 \frac{|d|}{\max_{d' \in L} |d'|}\right) (f(x) = \log_{10}(1+x))$$

where the TFIDF(d, t) is the TF-IDF score for the tuple of the material d and the word t as a query, L is a set of all the materials, and  $c_1$  and  $c_2$  are constants. This formula can be

interpreted as the difficulty of each word and its significance in a particular document contributing to the overall difficulty of the document. Besides, since the number of words composing the material is considered to affect its difficulty (the more words it includes, the more difficult it is), the value obtained by normalizing it is also used. Thus, the difficulty of the document will be computed as a total of products of the difficulty of the word and its significance.

#### Learner's difficulty preference estimation

Learners generally use English language materials of a difficulty level appropriate to their EFL proficiency. Previous research has confirmed a linear relationship between reading proficiency and the percentage of vocabulary in English language materials (Schmitt et al., 2011). As mentioned earlier, it is challenging to determine proficiency from learners' use of English language materials. However, this fact suggests that the difficulty of the materials used by learners is a good indicator of their EFL proficiency. In this study, learners' use of the learning materials was employed to estimate their preferences for the difficulty level of the materials.

According to the difficulty of the materials mentioned above, we extracted the top five most difficult materials from those used by the learners. We believe the difficulty of these five materials will best reflect the learner's difficulty preference because learners are expected to change the materials they use as their EFL proficiency improves. Then, the average difficulty of these five materials is calculated as the learner's difficulty preference. From now on, the difficulty level preferred by the learner *s* will be denoted as P(s). Namely, by denoting the five most difficult materials as  $d_1, d_2, d_3, d_4$ , and  $d_5, P(s)$  can be calculated as follows.

$$P(s) = \frac{1}{5} \sum_{i=1}^{5} D(d_i)$$

This feature guarantees that the value of P(s) is close to the maximum value of the learner's estimated difficulty preference.

# **Recommendation generation**

# Explanation for the recommendation

Our proposed recommendation algorithm is designed to provide feedback to the learner on the recommended material and the reason for recommending it. This allows the learner to be convinced of the recommendation and motivated to learn. This supplementary information includes the learner's difficulty preference (it implicates his/her proficiency), the recommended material's weight, and a sentence explaining why the material suits the learner.

D(d) - P(s)	Explanatory Sentence
less than -0.3 (very easy)	"This book is easy, but you can learn basic vocabulary with fun from this."
-0.3 to -0.1 (easy)	"This book is a little easy, but you can learn important vocabulary with this book."
-0.1 to 0.1 (average)	"This book is perfect for your English skills!"
0.1 to 0.3 (difficult)	"This book is a little difficult but worth trying!"
more than 0.3 (verv difficult)	"This book is challenging. Let's give it a try!"

**Table 1** Sentences that explain why the materials were recommended. They depend on the difference between the difficulty of the material and the learner's difficulty preference

We prepared five types of sentences explaining why the materials were recommended to the learner. According to the theory of proximal learning, when learners decide whether to study, they depend on their belief whether they already know the items. They will choose not to study if they believe they know the item already, and vice versa (Metcalfe & Kornell, 2005). Thus, the explanatory sentences should lead the learners to select the recommended materials. The types of sentences depend on the difference between the difficulty of the recommended material and the learner's preferred English difficulty level, i.e., the value of D(d) - P(s). This takes a value close to 0 when the difficulty of the recommended material d (i.e., D(d)) is close to the learner s's difficulty preference. When the value is smaller than 0, the recommended material is more accessible than the learner's level, and when it is larger than 0, the material is more complicated. The recommender provides different sentences according to the learner's difficulty with recommended materials, as shown in Table 1.

These messages were written to motivate learners to use the recommended materials. When the difficulty of the materials is close to the learner's estimated preference, the explanation states that this material perfectly fits the learner's English level. Besides, even if the recommended material is too easy or difficult, the recommender explains they are worth using.

#### **Recommendation weight**

The recommendation of materials by difficulty level refers to the recommendation of materials whose difficulty level is closest to the learner's estimated English level or the preference for the difficulty level of the material. In other words, recommended materials should not be too easy or too difficult because materials that are too easy or too difficult will negatively affect learning effectiveness. In this study, we introduce a recommendation weight for material *d* (denoted as R(s, d)), i.e., a value that indicates how much we recommend material *d* to learner *s*. We therefore define this value as follows:

$$R(s,d) = \log_{10} \left| \frac{1}{D(d) - P(s)} \right|$$

This formula takes a significant value if the difficulty level of the material d (i.e., D(d)) is close to the English preference level of the learner s (i.e., P(s)), and a small value otherwise. If the recommendation weight R(s, d) takes a significant value, it will recommend material d more strongly to learner s. This guarantees that the learner s receives the recommendation of the material d, whose difficulty is close to s's English skill levels. In the implementation, the value of D(d) - P(s) is adjusted so that it is not zero. The recommendation weights provided are linearly normalized so that the minimum value is 0.0 and the maximum value is 100.0 for ease of understanding by the user. The learner's preference and the difficulty level of the material are similarly normalized.

# Experiment for RQ1: Automated material difficulty estimation

#### Method overview

The difficulty estimation of English materials presented in this paper is essential for presenting how strongly recommended the recommended materials are for each learner. To show that this difficulty estimation mechanism can make accurate estimates, we conducted a simple experiment to measure the difficulty level of English materials. In this experiment, we used English picture books with difficulty levels ranging from A1 to B2/C1 on the CEFR level. Since these difficulty levels were displayed in the picture books in advance, we were able to check whether the formula presented above (D(d)) worked by comparing these difficulty level indicators with the difficulty level values estimated by the formula. Therefore, we calculated the difficulty level of the material busing the formula and summarized it using the difficulty level labels. The constants  $c_1$  and  $c_2$  were set to 0.1 and 1, respectively.

# Result

Table 2 shows the average number of books and estimated difficulty for each difficulty label. The results show that as the difficulty level of the label increases, the estimated difficulty level also increases. To confirm that the rank order of the estimated difficulty level coincided with the rank order of the label difficulty level, Kendall's rank correlation coefficient between the two was calculated, with a statistically significant value of 0.9820 (p < 0.001). This means that the order of difficulty labeled in the books and calculated in the formula has a strong positive correlation. Therefore, it is concluded that the formulas presented in this study can correctly estimate the difficulty level of the materials based on these characteristics.

Difficulty	# of books	Average of Estimated difficulty	
A1	193	0.857	
A1/A2	10	1.196	
A2	70	1.210	
A2/B1	92	1.210	
B1	10	1.743	
B1/B2	10	1.807	
B2	10	1.908	
B2/C1	10	1.995	

Table 2 Results of an evaluation of an equation that computes the difficulty of books

Focusing on the estimated difficulty level, it can be read that the increase in the calculated difficulty level with increasing difficulty level is significant for the materials with labeled difficulty levels from A1 to B1, but the increase in the estimated difficulty level with increasing difficulty level is slight for the materials labeled more difficult than B1 compared to the previous. This indicates that the increase in the difficulty level of the overall teaching material is smaller than that of the previous one. This may indicate that the effect of vocabulary difficulty on the overall difficulty of the materials is significant for the materials with low difficulty but decreases as the difficulty level increases.

# **Experiment for RQ2**

# **Method overview**

This experiment was conducted as part of the GOAL project to support students' independent learning. The experiment aimed to assess the extent to which recommendations of explainable materials were accepted by learners based on their behavior and perceptions, whether the recommendations influenced their learning behavior, and the characteristics of learners according to their perception of the recommendations.

# Setting

# Participants and experimental settings

This experiment was conducted in a Japanese junior high school with an online learning environment through Moodle (Moodle.org, 2017). In this school, all students had tablet PCs and access to the Internet at home, which allowed us to track and analyze students' learning logs in real time. The experiment involved 120 first-year students and 240 secondyear students. The explainable recommendation system was made available to all participants, and all were able to use the recommendations at any time during the experiment. Because junior high school is compulsory education in Japan, we did not conduct a multiple-condition comparison experiment to ensure equal educational opportunities. The first-year middle school students began participating on June 8, 2022,

Table 3 Summary of participants and experimental settings

Grade	Number of participants	Start date	End date	Number of days
1 <sup>st</sup> J.H.S.	120	June 8, 2022	July 20, 2022	43
2 <sup>nd</sup> J.H.S.	120	May 10, 2022	July 20, 2022	72

and the 120 second-year students began participating on May 10, 2022. The experiment was completed on July 20, just before summer vacation for first- and second-year students. A summary of the experimental setting is shown in Table 3.

#### Materials

When the experiment was conducted, 534 picture books for the ER program were stored in the library and categorized according to the difficulty level of the CEFR. Participating students were free to choose and read at their leisure. Table 4 shows the books in the library classified by CEFR level. The most common level was A1 or A1+ (for beginners, 205 books), followed by A2/B1 (for elementary and intermediate students, 126 books). Five books in the N/A level had no label regarding their level.

#### Evaluation

The data obtained in this experiment were evaluated from four perspectives: (1) acceptance of recommendations according to learning behavior, (2) acceptance of recommendations according to learner perceptions, (3) impact of recommendations on learner learning activities, and (4) characteristics of recommendations according to learner perceptions.

The evaluation phase used usage logs obtained from the BookRoll and GOAL systems, and a post-poll was conducted after the end of the ER program period. The usage logs were used to evaluate learning behaviors and learning skills. Three indicators used in the experiment were used to evaluate learning behavior: number of pages read by the learners, reading time, and number of words. Learning ability was assessed by the learners' reading speed (words per minute). The results of the post-poll were used to assess learners' perceptions of the recommendations.

Difficulty	# of books	
pre-A1	27	
A1 or A1+	205	
A1/A2	10	
A2 or A2+	75	
A2/B1	126	
B1 or B1+	37	
B1/B2	27	
B2	12	
B2/C1	10	
N/A	5	
Total	534	

Table 4 The number of picture books classified by CEFR levels

TAM in e-learning		Question item
Perceived usefulness	PU1.	The picture book recommendation has improved my learning performance.
	PU2.	The picture book recommendation has improved my learning efficiency.
	PU3.	The picture book recommendation helps me achieve a high level of learning.
Perceived ease of use	PEU1.	Using the picture book recommendation would make it easy to choose picture books to read.
	PEU2.	It is easy to learn how to use the picture book recommendation.
	PEU3.	It is easy to use the picture book recommendation.
Attitude	AT1.	It is good to read books using the picture book recommendation.
	AT2.	I would like to do learning activities with the picture book recommendation again.
	AT3.	I would like to use the picture book recommendation proactively.

**Table 5** List of questions in the post-poll and their class in TAM. Each question item was rated on a five-point Likert scale

# Post-poll

After the ER program was completed, a post-event survey on the recommender system was conducted. The poll was based on the Technology Acceptance Model (TAM) (Park et al., 2012) and aimed to investigate participants' perceptions of the system's usefulness, ease of use, and attitudes. TAM is an indicator commonly used to measure to what extent the system is accepted by users in the context of learning/educational support by information technologies (Granić & Marangunić, 2019). Each perception included three question items, for a total of nine items. All question items were rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A summary of the poll questions and their classification in the TAM is presented in Table 5.

#### Result

### System usage and recommendation acceptance

First, following the method of Takami et al. (2023), we define the terms "access" and "click" as used in this paper. We define "access" as an action in which a learner accesses the GOAL system's educational material recommendation page, and the system displays the five recommended educational materials. We also define "click" as an action in which a learner clicks on the title of a recommended teaching material and jumps to the book roll page. For example, if a learner accesses the GOAL system's recommended educational materials, these actions are counted as "one access and three clicks."



The number of accesses and clicks to the recommended picture book titles were counted to determine the frequency of use of the recommendation system. Figure 4 shows the number of accesses and clicks per day on the recommendation page, separately for first graders (left) and second graders (right). Total Clicks in the figure show the number of times an ER book was opened in BookRoll and whether a recommendation was used. Of the 240 participants, 100 students accessed the recommendation page at least once, and 46 students clicked on the title of the recommended material at least once. Figure 4 shows that except for a big spike in mid-June, the frequency of use of the recommendations was not very high. This result may reflect that all participants were introduced to the recommendation function by their teachers and tried using it in mid-June. Furthermore, first-year students used the ER materials without recommendations more frequently during the entire duration of the experiment. Other than that, there also was a small spike just before the end of the experimental period. In contrast, second-year students did not use ER materials without recommendations, except for big spikes in accesses in mid-May and mid-June. The first-year and second-year students had spikes just after the start of the experimental period in common. Table 6, which shows the daily averages of accesses and clicks, shows that participants accessed and clicked approximately 4 and 3 times daily, respectively. It also shows that first-year students used the recommendations more frequently than second-year students.

	# of accesses	# of clicks
1st Grade	3.881	3.881
2nd Grade	1.514	0.543
All	3.843	2.871

Table 6 Daily means of the number of accesses and clicks

Acceptance rates based on the number of accesses and clicks were also calculated to evaluate the degree to which the recommendations were accepted. First, the acceptance rate based on the number of accesses was calculated as the ratio of accesses in which one of the titles was clicked to all accesses. For example, suppose a learner accessed the recommendation page three times: on the first access, he/she clicked on three of the recommendation titles; on the second access, he/she clicked on one of the titles; and on the third access, he/she clicked on none of the titles. Then, the learner acceptance rate based on the number of accesses would be 66.7% (=2/3) since none of the titles were clicked twice out of the three accesses.

The acceptance rate based on the number of clicks was then calculated as the ratio of the number of recommended titles clicked on to the total number of recommended titles. In other words, in the above example, the learner clicked on four titles and a total of 15 (=3 x 5) recommended titles, resulting in a learner acceptance rate of 26.7% (=4/15) based on clicks.

The overall pass rate and the pass rate for each candidate were calculated in terms of both accesses and clicks. The results are shown in Table 7. This table shows that more than 40% of the accesses led to clicks on recommended books. However, the average acceptance rate for each student was lower than the overall average, indicating that not many students used the recommended books frequently. This implies that many students used the recommended books, but only a few used them continuously and repeatedly. The fact that the number of clicks was lower than the number of accesses also suggests that not many students clicked on more than one title in a single recommendation.

### Acceptance by learners' perception

The post-poll results were tabulated for the nine main items shown in Table 5. A total of 203 respondents responded to the posterior poll. Cronbach's alpha was calculated to be 0.884 (>0.8), indicating that the results of the posterior survey were internally consistent and reliable. Table 8 presents the results of the posterior survey by question item.

Table 8 shows that the respondents gave high ratings for ease of use (PEU) and attitude (AT) but relatively low ratings for usefulness (PU). Specifically, question AT1, which asked for a general opinion on whether using recommendations to read books was a good idea, had a high score. In contrast, questions PU1 and PU2, which asked how effective and efficient the recommendations were for learning, respectively, had a low score. Although

**Table 7** Acceptance rate of the recommendation. The value was calculated by access to the recommendation or the number of clicks on the title(s) of the recommended books

	Overall	Student mean	
By access	43.49% (=117/269)	26.66%	
By clicks	14.94% (=201/(269*5))	10.86%	

	PU1	PU2	PU3	PEU1	PEU2	PEU3	AT1	AT2	AT3
1	24	23	13	17	9	7	6	16	16
2	39	39	19	16	18	31	5	25	24
3	102	104	78	68	74	80	58	83	70
4	31	31	69	73	75	60	83	57	61
5	7	6	24	29	27	25	51	22	32
Mean	2.793	2.793	3.354	3.399	3.458	3.320	3.828	3.217	3.340

**Table 8** Summary of the post-poll results. This shows the number of participants who answered each question as 1 (strongly disagree) to 5 (strongly agree) and the mean values of the answers for each question

the learners did not feel that the recommendations helped them improve their EFL skills, the results suggest that the recommendations were not complex for them to use. Many learners accepted them favorably and thought using the recommendations was an excellent way to learn English. This suggests that many learners accept recommendations favorably and think that using recommendations to learn English is good. In other words, the use of recommendations was effective in motivating EFL learners.

# Learners' characteristics based on their perceptions

To examine learners' attitudes toward and characteristics of recommendations, Spearman's correlations between the learners' learning activities and indicators of their use of recommendations (number of pages read, time spent reading, speed of reading, number of words read, and acceptance rate by number of accesses and clicks) and the results of the posterior survey were calculated. Table 9 shows these correlations, which reveal that learners' awareness of the recommendations is significantly and slightly positively correlated with the number of pages, time, and words read, indicators of learning activity. Acceptance, however, showed little correlation with learners' awareness of recommendations, regardless of the number of accesses or clicks. This result suggests that learners who were highly motivated to engage in reading activities had a more positive attitude toward using recommendations than those who were not, which is consistent with the inferences made by the previous results.

Table 9 Correlation between learners' activities and poll answers

	PU1	PU2	PU3	PEU1	PEU2	PEU3	AT1	AT2	AT3
Pages	.120	.136	.136	.180**	.175*	.150*	.236***	.242***	.254***
Time	.105	.097	.195**	.182**	.097	.112	.252***	.282***	.260***
Speed	.094	.084	.001	.076	.125	.149*	.127	.099	.121
Words	.118	.125	.155*	.195**	.149*	.134	.263***	.281***	.290***
Accept (access)	.008	.060	.064	.032	.128	.140*	.093	.112	.120
Accept (clicks)	.009	.070	.059	.042	.128	.145*	.105	.126	.122
*	01 ***								

\*: *p* < .05, \*\*: *p* < .01, \*\*\*: *p* < .005

	Cluster 1	Cluster 2	Cluster 3
PU1	1.694 (0.786)	2.722 (0.639)	3.593 (0.812)
PU2	1.722 (0.779)	2.769 (0.635)	3.492 (0.858)
PU3	2.139 (0.961)	3.370 (0.705)	4.068 (0.848)
PEU1	1.806 (0.889)	3.519 (0.730)	4.153 (0.715)
PEU2	2.861 (1.334)	3.278 (0.747)	4.153 (0.665)
PEU3	2.861 (1.334)	3.065 (0.714)	4.068 (0.763)
AT1	2.889 (1.190)	3.731 (0.650)	4.576 (0.532)
AT2	1.778 (0.832)	3.148 (0.609)	4.220 (0.671)
AT3	1.889 (0.979)	3.259 (0.741)	4.373 (0.613)

**Table 10** The clustering result using the poll answers (mean (std.)). It also shows the breakdown of answers and learners' activities by cluster

For a more detailed analysis, the participating students were clustered using their responses to the post-event poll, and the mean of each indicator was calculated. The clustering algorithm used was K-means clustering. Table 10 presents the results of this analysis. First, the number of clusters was determined to be three using the elbow method (see Figure 5). The clustering results showed that students in Cluster 1 were the most negative about the recommendation, students in Cluster 3 were the most positive, and students in Cluster 2 were in between. Figure 6 shows that the values for each indicator are larger for clusters 1, 2, and 3, in that order, except for the pass rate by access. The table shows that students with more positive perceptions are more active in reading activities and read recommended books more often than those who do not. However, the acceptance rate by number of accesses was highest in Cluster 1, slightly lower than in Cluster 3, and lowest in Cluster 2. The fact that students with negative perceptions showed the highest acceptance rate by the number of accesses suggests that the need for recommendations is high but less effective in motivating learners to read. Students in Cluster 3 were probably more engaged in reading prior to the experiment, and the high click rate suggests that they were technologically familiar and found the multiple book recommendations helpful.





# Discussion

# RQ1: To what extent does the proposed mechanism for estimating the difficulty level of English materials for ER estimate the difficulty level of the materials?

The estimation results of the difficulty level of English materials showed that for relatively easy materials, the vocabulary level included tends to affect the overall difficulty significantly. In contrast, the effect of vocabulary is less significant for hard materials. As examples of the components of the difficulty of EFL materials, unfamiliar vocabulary (Kasim & Raisha, 2017; Qarqez & Ab Rashid, 2017) and unfamiliar content and grammatical matters (Satriani, 2018) can be raised. Considering that vocabulary and grammar are crucial in EFL learning, among these factors, the difficulty of grammar and content, rather than vocabulary itself, has a more substantial influence on the difficulty of the material in complex materials. Therefore, in response to the RQ, the present difficulty estimation system can accurately measure the difficulty of materials that are easier than those that are more difficult to some extent.

Although the difficulty estimation was conducted for extensive reading EFL materials in this study, the results can be generalized to other contexts if the materials and books are mainly composed of English texts. This difficulty estimation mechanism is entirely dependent on the content of the materials and is not related to the intended purpose of the materials. Therefore, measuring the difficulty of commercially available materials and texts written by EFL teachers, for example, is possible. However, it should be noted that this estimation is based on the CEFR-J, which, as mentioned earlier, is vocabularydependent and a difficulty level for EFL learners. In this study, the difficulty estimation was based on materials with pre-labeled difficulty levels. However, this difficulty estimation mechanism also allows us to recommend EFL materials that do not have pre-labeled difficulty levels.

However, there is room for improvement in the method used to generate explanations for recommendations based on the difficulty level of the material. In our work, we arbitrarily set the ranges that define which explanatory sentence is provided to learners. Therefore, it is difficult to evaluate recommendations quantitatively without a learner's subjective evaluation of the recommendation. We require a more detailed investigation to verify whether these values can really serve as a criterion to define a range of difficulties.

# RQ2: What characteristics of learners supported the recommendation of English language materials for ER that could be explained?

The results of the field experiment in the environment for education showed that few students used the recommendations repeatedly or continuously. On the other hand, students who used the recommendations were more active in their learning activities, suggesting a positive correlation between students' learning activities and the use of the recommendations. It also suggests that many students considered using recommendations a good thing and had a favorable impression, especially regarding ease of use. This idea is supported by the fact that the frequency of use increased only when the teacher introduced the recommendations to the students. Furthermore, although small, students who were highly motivated to read used the recommendations repeatedly, clicking on multiple titles from a single recommendation. Thus, although the recommendations did not motivate much learning, they were attractive, especially to highly motivated students to read.

Comparing the use of the recommendation system by participating first-year and secondyear students, more first-year students accessed the recommendations than second-year students. This can be attributed to the ability of the system and the learning with the system itself to arouse the learners' curiosity; the e-learning system itself has the power to arouse the curiosity of the learners themselves (Sarac et al., 2022). The learners' curiosity can explain the small spikes just after the start of the experimental period; they may be curious about the newly introduced function and tried using it. Besides, since first-year junior high school students who have just entered the school are unfamiliar with e-learning systems, it can be assumed that they were more interested in learning with the system and engaged in more learning activities. However, the effect of the system did not last long and did not arouse much curiosity in the second graders.

Other than that, we could find the slight increase in the number of accesses by the firstyear students just before the end of the experimental period. Since the end of the experiment coincides with the beginning of the summer vacation, it is thought that the teachers' announcements about learning during the summer vacation period increased the number of accesses and clicks.

#### Suggestion for the improvement of the recommendation

The results indicate that the recommendation of explainable English language materials was favorably accepted, especially by learners who were highly motivated to learn but were ineffective in motivating learners to learn. This suggests that the explanations for the recommendations, which were intended to motivate learners to learn, were not convincing enough to explain the rationale for the recommendations to the learners fully. The recommendation is based on the difference between the difficulty level of the material and the learner's English level, and the explanation is based on that difference. However, the explanation consists of a single sentence, which may be too brief for learners to understand why the recommendation was made entirely. Therefore, it is necessary to devise more detailed, precise, and persuasive explanations. For example, information on learners' learning behavior can be utilized since much data is accumulated in learning journals. Information such as how much of a specific material the learner has read or whether there are pages the learner reads repeatedly could be utilized. Since we currently only use data on vocabulary read by learners, more detailed information on learning logs would increase the detail and accuracy of the explanations.

We also noted that the effectiveness of the recommendations on learners' English learning was low. We speculate that this is due to the short duration of the experiment and the lack of intensive learning activities, suggesting that more experimentation is needed to introduce recommendations in more types of contexts other than ER. The Explainable English Material Recommendation is designed to adapt to various EFL learning contexts. For example, it can recommend English textbooks, reference books, or vocabulary and grammar quizzes. The learning benefits of this recommendation may become apparent in contexts where learners intensively study English.

# **Conclusion and future work**

This paper introduced an explainable English material recommendation targeting EFL learning. This recommendation is aimed at personalized English learning based on the learner's material preferences for the difficulty and the difficulty level of the material. Explainable English material recommendations recommend suitable English materials for learners based on their preferences and the difficulty level of the materials. To realize this feature of recommendation, we developed a mechanism to estimate the difficulty level of EFL materials automatically by employing TF-IDF, one of the information retrieval technologies. Using this mechanism, we calculated the difficulty level of EFL materials for extensive reading and showed that it can correctly measure the difficulty level. In particular,

the mechanism could accurately estimate the difficulty of the material, especially for relatively simple material where the influence of vocabulary on difficulty was significant. Then, we explained the details of the recommendation platform and its mechanism and experimented with 240 junior high school students. The experiment results showed that the recommendations did not affect students' learning motivation. However, they were favorably accepted by students who were highly motivated to learn in the first place, and their use was significantly and positively correlated with students' learning behavior. This indicated that the recommendation system was not effective in motivating learners in general but was readily accepted by naturally motivated learners.

There are two challenges to improving the recommendations and testing their effectiveness: further use of learning data and experimentation in various contexts. In this recommendation, the estimation of the difficulty of the material was based solely on vocabulary. The learner's estimated preference for English skill level was the average difficulty of the five most difficult materials that the learners had read in the past. Thus, the learning data used in this recommendation ignored detailed features of the language, such as English grammar and word usage, and learners' proficiency in each vocabulary and grammatical item. Therefore, the accuracy of the recommendation can be improved by utilizing further detailed learning data. In addition, more detailed and persuasive explanations that go into the learner's skill level can be provided by utilizing these data. This is expected to improve the system's effectiveness in motivating learners to learn.

Future directions in this research include improving the quality and explainability of recommendations by introducing a knowledge model. The only domain knowledge information used in this study was lexical profiles. However, it would be possible to make recommendations that go into the semantic content of the material, for example, using the semantic associations between words stored in the lexical knowledge map developed by Flanagan et al. (2019). Second, while our study did not focus on the learner's preference and needs, they are still important elements to make appropriate recommendation that can motivate the learner. Our work should aim to consider these elements to enable better learning outcomes resulting from the recommendation. In addition, we will conduct an experiment on the difference in the effect of the presence or absence of explanatory text in the recommendation on the learning effect as part of the future work. Another direction is the use of learner models that store information about the learner's skills. If we can establish a complementary relationship in which the recommendation system uses the learner model's performance information and, conversely, reflects the use of recommendations back to the learner model, we can expect to improve the quality of both the recommendation system and the learner model.

#### Abbreviations

AT: Attitude; BKT: Bayesian Knowledge Tracing; CEFR: Common European Framework of Reference for Languages; EFL: English as a Foreign Language; ER: Extensive Reading; EXAIT: Educational eXplainable Artificial Intelligent Tools; GOAL: Goal-Oriented Active Learning; IRT: Item Response Theory; LRS: Learning Record Store; PEU: Ease of Use; PU: Usefulness; SDL: Self-Directed Learning; TAM: Technology Acceptance Model.

#### Acknowledgements

The data reported in this paper and preliminary analysis has previously been presented in Takii et al. (2022):

Takii, K., Flanagan, B., Li, H., Yang, Y., & Ogata, H. (2022). Explainable English material recommendation using an information retrieval technique for EFL learning. In S. Iyer et al. (Eds.), *Proceedings of 30th International Conference on Computers in Education* (pp. 561–570). Asia-Pacific Society for Computers in Education.

#### Authors' contributions

KT, BF, HL, YY, KK, and HO contributed to the research conceptualization and methodology. Funding acquisition and supervision were done by HO. Software development, experiments, and data collection were performed by KT, HL, and YY. KT analyzed the data and wrote the manuscript. BF, HL, YY, KK, and HO provided comments to improve the manuscript. All authors read and approved the final manuscript.

#### Authors' information

Kensuke Takii is a Ph.D. student at the Graduate School of Informatics, Kyoto University. He received a bachelor's degree and master's degrees from the Graduate School of Informatics, Kyoto University. His research interests include: Learning Analytics, Educational Data Mining, Learner Modeling.

Brendan Flanagan is an Associate Professor at the Center for Innovative Research and Education in Data Science, Institute for Liberal Arts and Sciences, and the Graduate School of Informatics at Kyoto University. He received a bachelor's degree from RMIT University and master's and Ph.D. degrees from the Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests include: Learning Analytics, Educational Data Science, Educational Data Mining, NLP/Text Mining, Machine Learning, Computer Assisted Language Learning, and the Application of Blockchain in Education.

Huiyong Li is an Assistant Professor at Research Institute for Information Technology, Kyushu University. He received Ph.D. degree in Informatics from Kyoto University. His research interests include: Learning Analytics, Self-Directed Learning, Artificial Intelligence in Education, and Human-Computer Interaction.

Yuanyuan Yang is a post-doctoral researcher at Research Center for Intelligent Education, Zhejiang Lab, China. She received a Ph.D. degree in Informatics from the Graduate School of Informatics, Kyoto University. Her research interests include: Learning Analytics, Self-Directed Learning, Learning Performance Prediction, Educational Data Mining.

Kento Koike is an Assistant Professor at Faculty of Engineering, Tokyo University of Science. He received a Ph.D. degree in Engineering from Graduate School of Engineering, Tokyo Polytechnic University. His research interests include: Intelligent Tutoring Systems, Cognitive Science and Knowledge Engineering.

Hiroaki Ogata is a full Professor at the Academic Center for Computing and Media Studies, Kyoto University. His research interests include: Learning Analytics, Evidence-Based Education, Educational Data Mining, Educational Data Science, Computer Supported Ubiquitous and Mobile Learning, and CSCL.

#### Funding

This work was supported by JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2123. This work was also supported by JSPS Grant-in-Aid for Scientific Research (B) 20H01722 and 23H01001, JSPS Grant-in-Aid for Scientific Research (Exploratory) 21K19824, JSPS Grant-in-Aid for Scientific Research (B) 22H03902, NEDO JPNP20006 and JPNP18013.

#### Availability of data and materials

Data analyzed in this paper contains private information and therefore will not be made available.

#### Declarations

#### **Competing interests**

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Graduate School of Informatics, Kyoto University, Kyoto, Japan.

<sup>2</sup> Center for Innovative Research and Education in Data Science, Kyoto University, Kyoto, Japan.

<sup>3</sup> Research Institute for Information Technology, Kyushu University, Japan.

<sup>4</sup> Research Center for Intelligent Education, Zhejiang Lab, China.

<sup>5</sup> Faculty of Engineering, Tokyo University of Science, Tokyo, Japan.

<sup>6</sup> Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan.

Received: 31 January 2024 Accepted: 2 August 2024

Published online: 1 January 2025 (Online First: 10 September 2024)

#### References

- Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020, March). Complementing educational recommender systems with open learner models. In C. Rensing, H. Drachsler, V. Kovanović, N. Pinkwart, M. Scheffel & K. Verbert (Eds.), *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 360–365). ACM. https://doi.org/10.1145/3375462.3375520
- Abou-Khalil, V., Flanagan, B., & Ogata, H. (2021, June). Personal vocabulary recommendation to support real life needs. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin & V. Dimitrova (Eds.), *Artificial Intelligence in Education. AIED 2021. Lecture Notes in Computer Science, vol 12749* (pp. 18–23). Springer, Cham. <u>https://doi.org/10.1007/978-3-030-78270-2\_3</u>

Advanced Distributed Learning Initiative. (2013). *Experience API (xAPI) Standard*. <u>https://adlnet.gov/projects/xapi/</u> Anand, R., & Jeffrey, D., U. (2011). *Mining of massive datasets*. Cambridge University Press.

- Bahrani, P., Minaei-Bidgoli, B., Parvin, H., Mirzarezaee, M., & Keshavarz, A. (2024). A hybrid semantic recommender system enriched with an imputation method. *Multimedia Tools and Applications*, 83(6), 15985–16018. <u>https://doi.org/10.1007/s11042-023-15258-4</u>
- Bourkoukou, O., & El Bachari, E. (2018). Toward a hybrid recommender system for e-learning personalization based on data mining techniques. *International Journal on Informatics Visualization*, 2(4), 271–278. <u>http://dx.doi.org/10.30630/joiv.2.4.158</u>
- Chen, C. N., Chen, S. C., Chen, S. H. E., & Wey, S. C. (2013). The effects of extensive reading via e-books on tertiary level EFL students' reading attitude, reading comprehension, and vocabulary. *Turkish Online Journal of Educational Technology*, 12(2), 303–312.
- Chen, W., Shen, Z., Pan, Y., Tan, K., & Wang, C. (2024). Applying machine learning algorithm to optimize personalized education recommendation system. *Journal of Theory and Practice of Engineering Science*, 4(01), 101–108. <u>https://doi.org/10.53469/itpes.2024.04(01).14</u>
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4, 253–278. <u>https://doi.org/10.1007/BF01099821</u>
- Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). https://www.coe.int/lang-CEFR
- Dai, Y., Flanagan, B., Takami, K., & Ogata, H. (2022, March). Design of a user-interpretable math quiz recommender system for Japanese high school students. In B. Flanagan, R. Majumdar, H. Li, A. Shimada, F. Okubo & H. Ogata (Eds.), Proceedings of the 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior (pp. 30–38). CEUR Workshop Proceedings.
- Day, R. R., Bamford, J., Renandya, W. A., Jacobs, G. M., & Yu, V. W. S. (1998). Extensive reading in the second language classroom. *RELC Journal*, 29(2), 187–191.
- Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, 52, 338–348. <u>https://doi.org/10.1016/i.chb.2015.05.041</u>
- Enisa, M., İnceçay, G., & Incecay, V. (2013). Fostering learner autonomy through extensive reading: The case of oral book reports. *ELT Research Journal*, 2(1), 16–25.
- Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. Knowledge Management & E-Learning, 10(4), 469–484. <u>https://doi.org/10.34105/j.kmel.2018.10.029</u>
- Flanagan, B., Chen, M. A., Lecailliez, L., Majumdar, R., Akçapinar, G., Ocheja, P., & Ogata, H. (2019, December). Automatic vocabulary study map generation by semantic context and learning material analysis. In M. Chang et al. (Eds.), *Proceedings of the 27th International Conference on Computers in Education* (pp. 698–702). Asia-Pacific Society for Computers in Education.
- Flanagan, B., Takami, K., Takii, K., Dai, Y., Majumdar, R., & Ogata, H. (2021). EXAIT: A symbiotic explanation learning system. In M. M. T. Rodrigo et al. (Eds.), *Proceedings of the 29th International Conference on Computers in Education Conference* (pp. 404–409). Asia-Pacific Society for Computers in Education.
- George, G., & Lal, A. M. (2019). Review of ontology-based recommender systems in e-learning. Computers & Education, 142, 103642. <u>https://doi.org/10.1016/j.compedu.2019.103642</u>
- Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. British Journal of Educational Technology, 50(5), 2572–2593. <u>https://doi.org/10.1111/bjet.12864</u>
- Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.
- Hsu, M. H. (2008). A personalized English learning recommender system for ESL students. *Expert Systems with Applications*, *34*(1), 683–688. <u>https://doi.org/10.1016/j.eswa.2006.10.004</u>

- Kasim, U., & Raisha, S. (2017). EFL students' reading comprehension problems: Linguistic and non-linguistic complexities. *English Education Journal*, 8(3), 308–321.
- Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, 56(3), 885–899. <u>https://doi.org/10.1016/j.compedu.2010.11.001</u>
- Li, H., Majumdar, R., Chen, M. R. A., & Ogata, H. (2021). Goal-oriented active learning (GOAL) system to promote reading engagement, self-directed learning behavior, and motivation in extensive reading. *Computers & Education*, 171, 104239. <u>https://doi.org/10.1016/j.compedu.2021.104239</u>
- Majumdar, R., Yang, Y. Y., Li, H., Akçapinar, G., Flanagan, B., & Ogata, H. (2018). GOAL: Supporting learner's development of self-direction skills using health and learning data. In J. C. Yang et al. (Eds.), *Proceedings of the* 26th International Conference on Computers in Education (pp. 406–415). Asia-Pacific Society for Computers in Education.
- Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 25(1), 91–102. https://doi.org/10.1016/S0346-251X(96)00063-2
- Mermelstein, A. D. (2015). Improving EFL learners' writing through enhanced extensive reading. *Reading in a Foreign Language*, 27(2), 182–198.
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52(4), 463–477. https://doi.org/10.1016/j.jml.2004.12.001
- Moodle.org. (2017). Moodle Open-source learning platform Moodle.org. https://moodle.org/
- Ningsih, S. (2019, March). Developing students level of autonomy through extensive reading activity in EFL context. In A. G. Abdullah, A. A. Danuwijaya, N. Haristiani, R. Dallyono, Y. Wirza & D. Yuliana (Eds.), *Proceedings of Second Conference on Language, Literature, Education, and Culture* (pp. 355–359). Atlantis Press. https://doi.org/10.2991/icollite-18.2019.78
- Ochoa, X., & Duval, E. (2009). Quantitative analysis of learning object repositories. *IEEE Transactions on Learning Technologies*, 2(3), 226–238. <u>https://doi.org/10.1109/TLT.2009.28</u>
- Ooge, J., Kato, S., & Verbert, K. (2022, March). Explaining recommendations in e-learning: Effects on adolescents' trust. In Proceedings of 27th International Conference on Intelligent User Interfaces (pp. 93–105). ACM. https://doi.org/10.1145/3490099.3511140
- Park, S. Y., Nam, M. W., & Cha, S. B. (2012). University students' behavioral intention to use mobile learning: Evaluating the technology acceptance model. *British Journal of Educational Technology*, 43(4), 592–605. <u>https://doi.org/10.1111/j.1467-8535.2011.01229.x</u>
- Pitts, M., White, H., & Krashen, S. (1989). Acquiring the second language vocabulary through reading: A replication of the clockwork orange study using second language acquirers. *Reading in a Foreign Language*, 5(2), 271–275.
- Qarqez, M., & Ab Rashid, R. (2017). Reading comprehension difficulties among EFL learners: The case of first and second year students at Yarmouk University in Jordan. *Arab World English Journal*, 8(3), 421–431. <u>https://dx.doi.org/10.24093/awej/vol8no3.27</u>
- Salehi, M., & Kmalabadi, I. N. (2012). A hybrid attribute–based recommender system for e-learning material recommendation. *IERI Procedia*, 2, 565–570. <u>https://doi.org/10.1016/j.ieri.2012.06.135</u>
- Sarac, S., Mede, E., & Akgun, E. (2022). Does e-learning trigger epistemic curiosity?. Journal of Qualitative Research in Education, 30. <u>https://doi.org/10.14689/enad.30.9</u>
- Satriani, E. (2018). Reading comprehension difficulties encountered by English students of Islamic University of Riau. J-SHMIC: Journal of English for Academic, 5(2), 15–26. <u>https://doi.org/10.25299/jshmic.2018.vol5(2).1885</u>
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The*
- Modern Language Journal, 95(1), 26–43. <u>https://doi.org/10.1111/j.1540-4781.2011.01146.x</u>
- Shishehchi, S., Banihashem, S. Y., Zin, N. A. M., & Noah, S. A. M. (2011, June). Review of personalized recommendation techniques for learners in e-learning systems. In S. A. M. Noah, N. Omar, F. Crestani, M. Mohd, M. J. A. Aziz & O. A. Puade (Eds.), *Proceedings of 2011 International Conference on Semantic Technology and Information Retrieval* (pp. 277–281). IEEE. <u>https://doi.org/10.1109/STAIR.2011.5995802</u>
- Sikka, R., Dhankhar, A., & Rana, C. (2012). A survey paper on e-learning recommender system. International Journal of Computer Applications, 47(9), 27–30. <u>https://doi.org/10.5120/7218-0024</u>
- Sun, Y. C. (2003). Extensive reading online: An overview and evaluation. *Journal of Computer Assisted Learning*, 19(4), 438–446. <u>https://doi.org/10.1046/i.0266-4909.2003.00048.x</u>
- Takahashi, W., & Umino, T. (2020). Out-of-class extensive reading in Japanese as a second language: Enhancing learner autonomy beyond the classroom. *Electronic Journal of Foreign Language Teaching*, 17(1), 50–63.
- Takami, K., Flanagan, B., Dai, Y., & Ogata, H. (2023). Personality-based tailored explainable recommendation for trustworthy smart learning system in the age of artificial intelligence. *Smart Learning Environments*, 10, 65. <u>https://doi.org/10.1186/s40561-023-00282-6</u>
- Takii, K., Flanagan, B., & Ogata, H. (2021). An English picture-book recommender system for extensive reading using vocabulary knowledge map. In M. Scheffel, N. Dowell, S. Joksimovic & G. Siemens (Eds.), Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge (pp. 40–42). ACM.
- Tanaka, H., & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading. *The Reading Matrix*, 7(1).

Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. Artificial Intelligence Review, 50, 21–48. <u>https://doi.org/10.1007/s10462-017-9539-5</u>

Tono, Y. (2020). The CEFR-J Wordlist Version 1.6. http://www.cefr-j.org/download.html

Truong, H. M. (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*, *55*, 1185–1193. <u>https://doi.org/10.1016/i.chb.2015.02.014</u>

Urquhart, A. H., & Weir, C. J. (2014). *Reading in a second language: Process, product and practice*. Routledge. xAPI.com. (2011). *What is an LRS? Learn more about Learning Record Stores*. <u>https://xapi.com/learning-record-store/</u>

Yang, Y., Majumdar, R., Li, H., Flanagan, B., & Ogata, H. (2024). Design of a learning dashboard to enhance reading outcomes and self-directed learning behaviors in out-of-class extensive reading. *Interactive Learning Environments*, 32(3), 892–909. https://doi.org/10.1080/10494820.2022.2101126

Zou, D., & Xie, H. (2018). Personalized word-learning based on technique feature analysis and learning analytics. Journal of Educational Technology & Society, 21(2), 233–244.

# **Publisher's Note**

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

*Research and Practice in Technology Enhanced Learning (RPTEL)* is an open-access journal and free of publication fee.