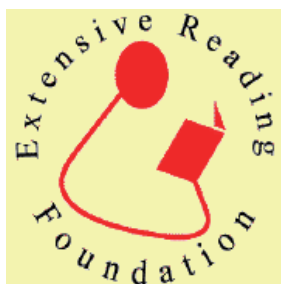


Developing an Online Graded Reader Database to Support ER



MARK BRIERLEY, RYOMA NAKAMURA AND
MASAAKI NIIMURA

Shinshu University

This is a progress report on the development of an online system to support Extensive Reading (ER) by keeping records of books which learners have read. Such a system requires data including word counts, levels of difficulty, and cover images, which must be collected and organised, and regularly updated as new books are published, or more information becomes available on existing books. There are significant costs involved in maintaining this data, both in terms of time and money, and some of this work is currently being duplicated unnecessarily by different ER programmes. Our aim is to develop a web service to maintain and make available such ER data either through an API (Application Programming Interface) that can be added to existing online systems, or through a dedicated website. This service will regularly collect and save data on graded readers from online sources, converting data into a standard format where necessary.

The essential components of an Extensive Reading (ER) programme are time, access to books, and permission to enjoy reading. In order for the large amounts of reading suggested by the term “extensive” to take place, and in view of the different levels and interests of students, a wide range of books is necessary, and information is needed on those books. Besides being able to consult a list of available titles, arranged by publisher, students and ER practitioners also need to know how easy each book is, how long it is, how good it is and which students will like it. This paper discusses the various types of information which are needed for ER programmes to be run successfully and the ways in which the database under development aims to meet those needs.

Difficulty Levels of Graded Readers

The reading level of a book is critical. Vocabulary research suggests that at least 98% of the words on any page must be known in order for fluent reading to be possible (Nation, 2001). In contrast to Stephen Krashen’s $i+1$ in the natural order hypothesis (1982), Richard Day has suggested students should be reading at $i-1$ (1998) and Atsuko Takase and Hitoshi Nishizawa have suggested that $i-2$ is better, with research showing that students who read easier books make more progress in language proficiency (Takase & Nishizawa, 2010). Reading level, however, is not straightforward and “ i ” is a notion rather than a precise linguistic unit. Krashen’s original hypothesis posits that language is acquired in a particular order, and if a learner has acquired grammatical structure i , they will next acquire item $i+1$. The particular order has not been defined, and it is not clear whether there is a universal order or the extent to which individual learners follow an order, so Krashen’s hypothesis remains controversial.

Brierley, M., Nakamura, R. and Niimura, M. (2020). Developing an online graded reader database to support ER. *Extensive Reading World Congress Proceedings*, 5, 219-228.

Publishers of graded readers each have their own levelling systems based on different word lists, inventories of grammatical structures, or else simply on the intuition of experienced editors. The EPER Scale from the Edinburgh Project on Extensive Reading was the first attempt to organize the graded readers from all publishers into a set of levels of difficulty, ranging from the easiest level (G) to the hardest (A), with a level X for books above that. More recently, Akio Furukawa, Minami Kanda and others (2013) developed the Yomiyasusa Level (YL; 読みやすさ literally meaning “reading ease”) which assigns a score from 0.1 to 10.0 on the basis of the impressions of readers in Japan. Since then Tom Robb has developed the Kyoto scale, which is used with MReader (n.d.). Under the auspices of the Extensive Reading Foundation, Rob Waring has developed the ERF Scale, which uses the number of headwords as the primary scale of difficulty (2016). The ERF scale began with 16 levels, later expanding to 20 to cover mid-range graded readers that bridge language learner literature and

books aimed at adult native speakers. In addition to these scales, several libraries use their own de facto levelling systems, for example by using colour-coded labels to indicate the difficulty of the books. Some scales are compared in Table 1, and Figure 1 shows an excerpt from the ERF Graded Reader Database. While the scales broadly correlate, it should be pointed out that a particular book may appear at different points on different scales since each scale considers different factors when assessing how easy a book is to read (see Gillis-Furutaka, 2015 for more factors that affect book difficulty). In addition to these explicit scales, we should mention research by Holster et al. (2017), who found that word count is the most reliable indicator of how difficult a student will find a book.

Graded Reader Word Counts

In addition to knowing how difficult a book will be to read, practitioners and participants in ER programmes may also want to know how many words are in each book

Table 1: Comparison of Different Reading Scales

Scale	Beginner				Elementary			Intermediate			Upper			Advanced		
ERF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
YL		0.4	0.7	0.8	1.0	1.2	1.4	1.8	2.2	3.1	3.8	4.6	5.3	6.0	6.8	7.6
Kyoto		S	1	2	3	4	5	6	7		8			9		
EPER				G	F	E		D			C	B	A	X		
Cambridge				S	1		2			3		4		5		6
Macmillan			1	2	3		4				5					
Oxford				S	1		2	3		4	5			6		

=HYPERLINK("https://sites.google.com/site/erfgrlist/", "Extensive Reading Foundation Graded Reader List")																		
Extensive Reading Foundation Graded Reader List																		
Atama-ii Books		Level Info							Word count / Length			M-Reader		Genre & Audience			Publication I	
ERF code	Title ERF Language Learner Literature Awards: LLLA Winners in Red, Finalists in Blue	Author(s)	Series	Publisher level	Headwords	ERF level (E) Early (M) Mid (H) High	Common European Framework CEF level	Yomi-yasusa level (YL) Average	Word count (body only)	Additional material wordcount	Actual or Estimated count?	Book length (Pages)	Quiz Available? (Y/yes, (N)/no, (S)/soon	MReader Reading Level (RL)	Fiction (F) Nonfiction (NF)	Genre mystery, adventure, love, horror, etc.	Target age group: children, teen, adult, etc.	ISBN A (print, no audio)
4	ATM Backstage Pass	Lesley Ito	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140345
5	ATM Brainstorm	Andy Boon	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140406
6	ATM Hunter in the Darkness	Ramy Habeeb	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140352
7	ATM Immortal Empire	Marcos Benevides	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140369
8	ATM Journey to Mars	Paul Raine	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140314
9	ATM Lost Cup, The	James Broadbridge	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140338
10	ATM North Fence, The	Theresa Kannenberg	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140383
11	ATM Ooze, The	Kyle MacLachlan	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140376
12	ATM Tomado Alley	Lesley Ito	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140390
13	ATM Zombies in Tokyo	Andy Boon	Atama-ii Books	1	300	E-M	A1+	0.9	2,500	1,000	estimated	50	Y	1	F	adventure, multi-path 8 endings	11+ (PG)	9781941140321

Figure 1. ERF graded reader data

since the “extensive” of ER refers to reading quantity. Keeping a record of the number of words that students read makes it possible to assess the effectiveness of reading. Such a record can also be used as an assessment tool in language classes. To calculate how much students have read, we need data on the number of words in each book.

A Prototype Online Support System for ER

The origins of this research project come from an online support system developed within a Japanese national university in 2008 to track word counts (ERS: Extensive Reading System; Brierley, Ruzicka, Sato, & Wakasugi, 2009; Wakasugi, Sato, Niimura, Brierley, & Kunimune, 2009; Brierley & Ruzicka, 2012). Before the online system, word counts were tracked through an increasingly complex system based on booklets of word counts and paper forms. Teachers maintained the word count booklets, adding or editing data as new books were added to the class libraries, and students referred to these when adding word counts to their book records.

Recent Versions of ERS

Until now, ERS has only been available to students within the university. The newest version will be available as a Moodle plug in. The data used by ERS will be available via API (Application Protocol Interface) and it will be possible to access this data through a dedicated website.

ERS allows students to search for a book online when they start reading, after which the selected book appears on their top page. When they indicate they have finished reading, they are asked how much of the book they read, since one of the principles of ER is that students should abandon books if they find them too difficult or uninteresting. There are also simple questions asking how interesting the book was and how easy it was, on scales from 1 (difficult or boring) to 5 (easy or interesting). The book is then added to a reading record and the number of words added onto the total word count. If students did not read a whole book, the word count is multiplied by the percentage they have read. As students read more, a graphic develops from a sprout to a seedling and eventually

The screenshot shows a user profile for '中村 涼真's "ER Tree' on a dark background. On the left is an illustration of a tree with blue leaves and a brown cat sitting at its base. On the right, there are two yellow buttons: 'Review' and 'Delete'. Below these is a 'Reading Status' section with the following information:

- 3866 words.
- 0 books.
- 60 hours.
- Last submit is 0000-00-00.

Record of Reading

Title	Pages	Words	YL
Buzz, Said the Bee	29	165	0.5
A Kidnapped Santa Claus	0	3700	4

Recommend

Figure 2. ERS 4.0 top page

Search Books

Search by Title

tom sawyer

- Adventure of Tom Sawyer and Adventure of Huckleber ISBN: 9780451528643
- Adventures of Tom Sawyer, The ISBN: 9780582419230
- Adventures of Tom Sawyer, The ISBN: 9784896843026
- Adventures of Tom Sawyer, The ISBN: 9780435273361
- Adventures of tom Sawyer, The ISBN: 9781405076081
- Adventures of Tom Sawyer, The ISBN: 9780194229364
- Adventures of Tom Sawyer, The ISBN: 9780194789004
- Adventures of Tom Sawyer, The ISBN: 9784896840438
- Adventures of Tom Sawyer, The ISBN: 9781405842778

Figure 3. Will the real Tom Sawyer please stand up!

Title	Adventures of Tom Sawyer, The		
ISBN	9780194229364		
Pages	41		
Words	5845		
YL			

How much of the book did you read?*

How long did it take?*

How interesting was it?*

How difficult was it?*

Was the story easy to understand?*

How many difficult words were there?*

Do you have any impressions?

Figure 4. Page for students to report on the book they have read

to a tree to indicate progress. Version 4 of ERS is currently under trial. Figure 2 shows the top page, Figure 3 shows a search page, and Figure 4 shows the page for students to report on the book they have read.

Updating and Expanding the Database

Maintaining an up-to-date database of books is a constant challenge for any online ER support system (see Brierley, M., Gillis-Furutaka, A., et al., 2019). The first version of ERS was developed with limited book data, so a function was added to allow students to add new books along with their word counts and YL. This data was then confirmed by a teacher or administrator. For a subsequent version, book data was obtained with the help of Akio Furukawa from SEG (Science Education Group 科学的教育グループ), and because this data appeared to be so comprehensive the function for students

to directly add books was not added to the updated system. However, publishers continued to bring out new books and series, and students continued to read books that were not included in the data held in the system, so an external form was created to register new books to the system. The information input to the form had to be checked and then the database edited within the system.

The latest ERS version includes forms for new books to be added. This is accessed when a search by name fails, and a subsequent request for the ISBN of the book is not matched by an entry in the database. Administrators are also able to edit book data. Figure 5 shows a search page where an unsuccessful search by name gives an option to search by ISBN. Figure 6 shows a form for adding new books and Figure 7 shows the administrator page.

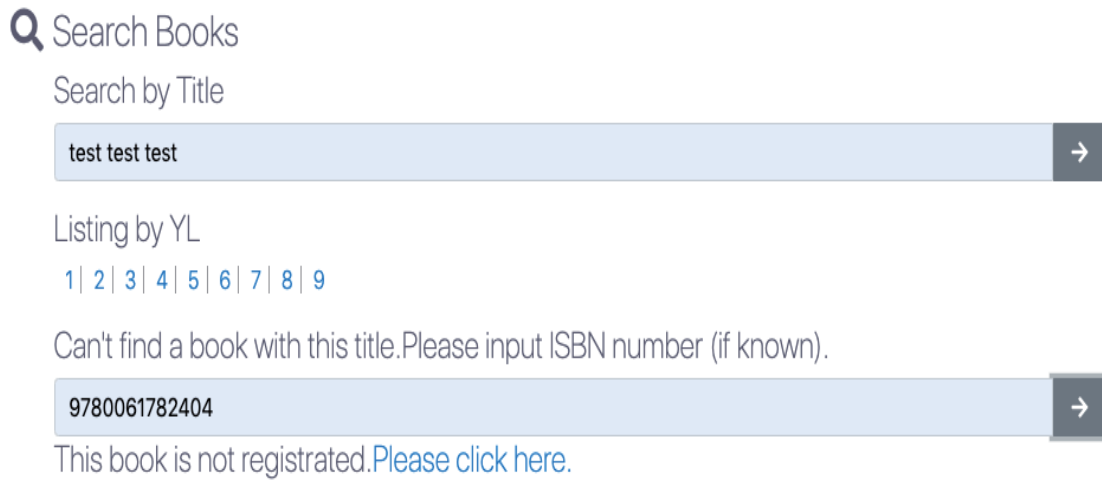


Figure 5. ERS unsuccessful search showing ISBN search option.

The screenshot shows a form titled 'Registration Book' with a plus icon. Below the title is a grey bar with the text 'Please input book info to form.'. The form is divided into two sections: 'Basic Information' and 'Further Info'. The 'Basic Information' section contains the following fields: ISBN* (input field with '9780061782404'), Title* (input text), Word count(body only)* (input number), Actual or Estimated count? (dropdown menu with a plus icon), YL (input number), Author (input text), Series (input text), Publisher Level (input number), and Book Length(pages) (input number). The 'Further Info' section contains a 'memo' field with a large text area.

Figure 6. ERS student registration page for new book

ISBN	Title	WordCount	ActualOrEstimatedCount	Author	Series	PublisherLevel	BookLength(Page)	He
9781416550556	The Forgotten Garden	182600						
9780746096611	Tortoise and the Eagle	244						
9780590441858	Buzz, Said the Bee	165						
9781405057783	Lis for Lawless	26000						
9781920407001	Lis for Lawless	26295						
0	Lis for Lawless	27331						
9780688147365	More More More, Said	360						
0	Simpsons Guide to Spr	0						
9780590259453	What is That?Said the	118						
9780887306310	Yes or No: The Guide to	23138						
9780684865744	'Tis	141000						
9780370328768	006 and a Bit	845						
9780688065225	1 hunter	43						

Figure 7. ERS administrator user interface for book data

Book Covers

Book cover images were displayed in the search area, and in students’ records of the books they had read. Cover images have two important functions: first, several titles are available from multiple publishers at various levels, and cover images help students identify the correct book; secondly, images of books added to a student’s pages provide a visual record of achievement similar to the collections of badges or stickers which have been used throughout their education.

Book images were initially accessed from amazon.com, where they were freely available. However, amazon.com changed its policies so that an account was required to access book cover image files. Some book images are available via Calil corporation’s OpenBD (n.d.), which is a comprehensive database of books in many Japanese libraries. However, cover images of many books with overseas ISBNs are missing. Recently

XReading and MReader have generously provided cover images and in the future we hope to make these images publicly available, and to add more images to the database as more books are published and more books being used for ER are discovered.

Student Data

Further to data on books, online support systems can add data on students. Such data may allow automatic recommendation of suitable books to students based on what they have read, and the reading diets of other students with similar tastes. In order to do this, large data sets are needed, and smaller institutions may not have sufficient students to provide meaningful data (see Figure 8). In order to meet this challenge we propose partial sharing of data from reading support systems of multiple universities as shown in Figure 9.

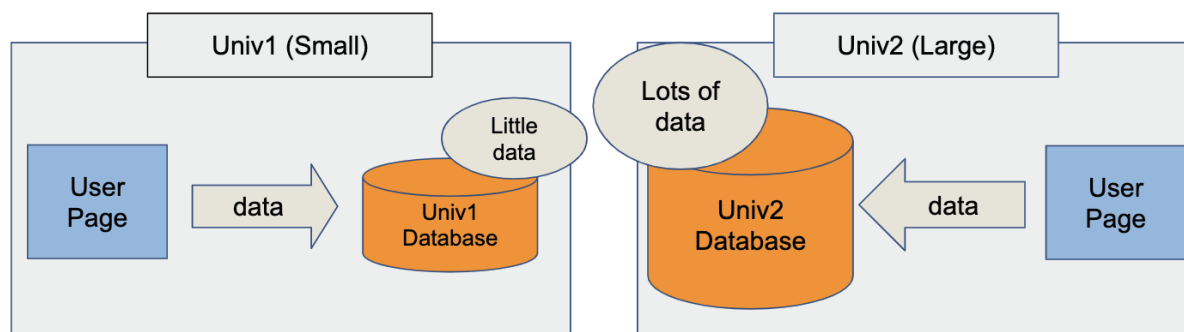


Figure 8. Challenges for a recommendation function

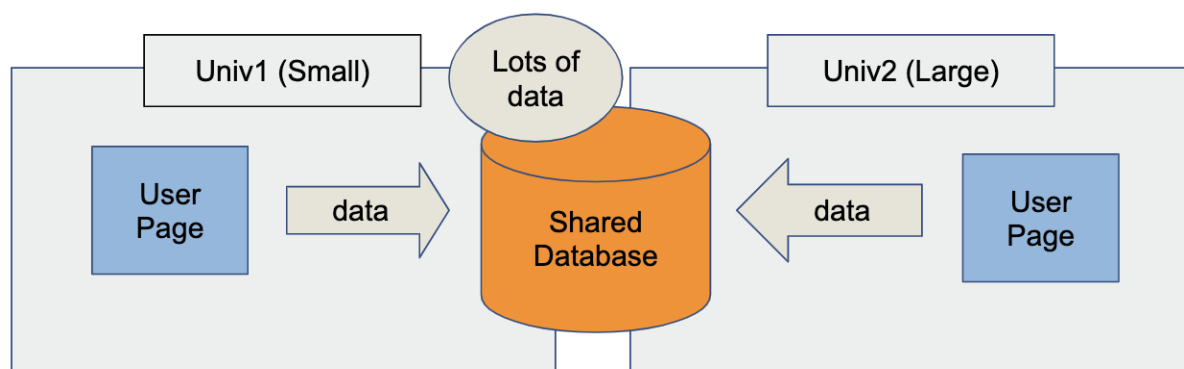


Figure 9. Database sharing

This system of sharing will more efficiently collect reviews and other data which could be used for a book recommendation function. In order to protect privacy, personal data can be removed from the book data. Data such as word count and YL are stored in a shared database. We are currently developing a system that allows administrators and other stakeholders to edit this data and add new data for new publications. A multi-layered database can be implemented, as in Figure 10.

Personal information for each institution can be kept in a database within each institution while a shared database is

accessible from each institution. When a user performs an operation from the page in their institution, first the local database will be accessed. If data is not available in the local database, a query will be sent to the shared database. An advantage of this system is that each educational institution only needs to create a user interface, and any systems using the shared database just need to program inquiries to meet the specified format. This should reduce development costs. The above system will be developed with careful consideration of personal data to ensure that only suitable information is shared.

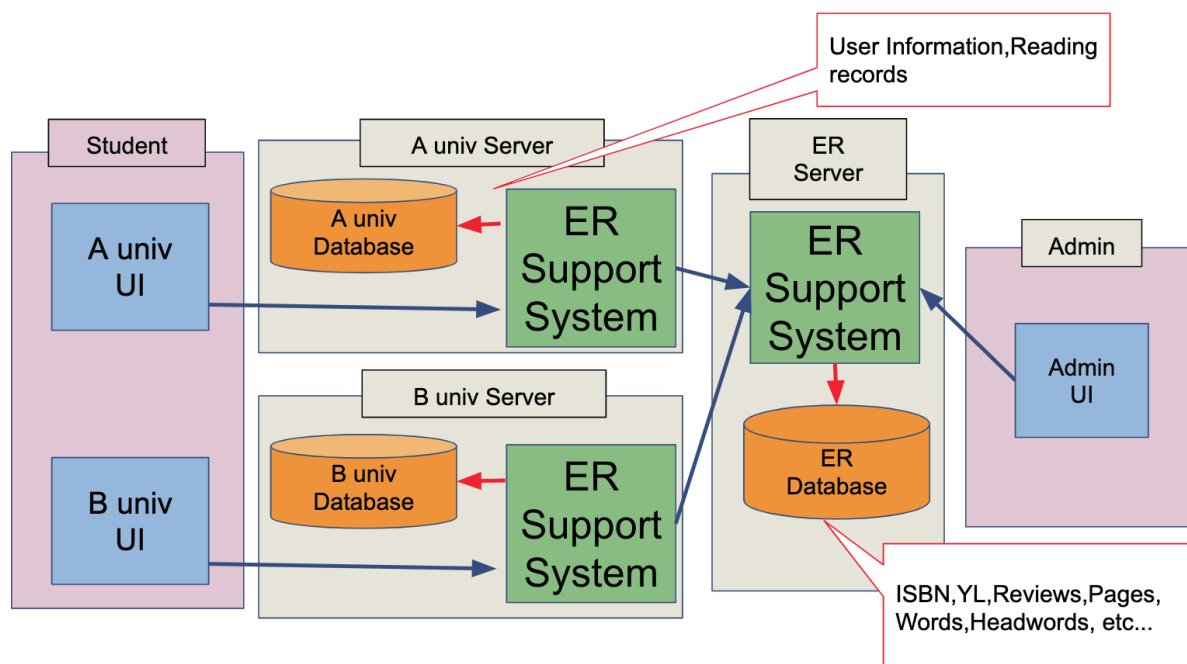


Figure 10. Multi-layered database

Conclusion

Collecting and maintaining data on graded readers is essential for successful ER programmes, but presents many challenges. This report introduces a new version of a book recording system that can be used in multiple institutions and will provide access to data on the level of the books, the number of words, their popularity among student readers, and provide recommendations for books that students will probably enjoy reading. We also show the organization of databases so that data may be shared between institutions while preserving the privacy of the users.

Acknowledgements

This research was supported by Japan Society for the Promotion of Science *Kakenhi* grant number JP17K01119.

References

- Brierley, M., Gillis-Furutaka, A., Niimura, M., Ruzicka, D., Takahashi, M., & Yoshioka, T. (2019). Towards a Graded Reader Database. *PanSIG 2018 Journal* (pp. 21–27). Tokyo: JALT. http://pansig.org/publications/2018/PanSIG_2018_Journal.pdf
- Brierley, M., & Ruzicka, D. (2012). Developing institutional relationships through ER. In K. Bradford Watts, R. Chartrand, & E. Skier (Eds.), *The 2011 Pan-SIG Conference Proceedings* (pp. 220-234). Matsumoto: JALT.
- Brierley, M., Ruzicka, D., Sato, H., & Wakasugi, T. (2009). An online system for assessing Extensive Reading. In Alan Stoke (Ed.) *JALT2008 Conference Proceedings* (pp. 1047–1057). Tokyo: JALT.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*.

- Cambridge: Cambridge University Press.
- Furukawa, A., Kanda, M., et al. (2013). *Eigo tadoku kanzen book guide [The Complete Book Guide for Extensive Reading]*. Tokyo: Cosmopier.
- Gillis-Furutaka, A. (2015). Graded reader readability: Some overlooked aspects. *Journal of Extensive Reading*, 3(1), 1–19. <http://jaltpublications.org/content/index.php/jer/article/view/7/1>
- Holster, T. A., Lake, J., & Pellowe, W. R. (2017). Measuring and predicting graded reader difficulty. *Reading in a Foreign Language*, 29(2), 218–244. <http://nflrc.hawaii.edu/rfl/October2017/articles/holster.pdf>
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. London, UK: Pergamon Press.
- MReader. (n.d.). *The Kyoto Scale*. <https://mreader.org>
- Nation, I. (2001). *Learning Vocabulary in Another Language*. Cambridge, England: Cambridge University Press.
- Open BD. (n.d.). <https://openbd.jp/>
- Takase, A., & Nishizawa, H. (2010). Two critical tips to motivate EFL learners to read extensively. In A. Harris & A. Brandt (Eds.), *Language, Learning, and Context. Proceedings of the 42nd Annual Meeting of the British Association for Applied Linguistics, 3-5 September 2009, Newcastle University* (pp. 135–137). London: Scitsiugnil Press.
- Wakasugi, T., Sato, H., Niimura, M., Brierley, M., & Kunimune, H. (2009). Developing and managing an online system to support extensive reading. In T. Bastiaens et al. (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 3746–3752). Chesapeake, VA: AACE.
- Waring, R. (2016). *The Extensive Reading Foundation Graded Reader Level Scale: Chart*. http://www.robwaring.org/er/scale/ERF_levels.htm
- Xreading. (n.d.) Retrieved from <https://www.xreading.com/>