



An evaluation of a Lexiles–based reading management system

Gloria Shu Mei Chwo^{1,*}

¹General Education Center, Hungkuang University, Taichung, Taiwan

Abstract

This study reports on a trial use, in independent extensive reading over a three–month period, of an online reading management system, with low proficiency non–native speakers of English (EFL/TESOL) at a Taiwanese university, in order to evaluate its effectiveness in increasing reading proficiency. The system used the controversial Lexile measure, of both text difficulty and reader ability level, to match readers to texts. Data was gathered with the software package’s objective measures of reading proficiency, text comprehension and number and type of texts read, to which the researchers added a reader attitude questionnaire. While student attitudes towards the visual aspects of the system and its wide range of reading topics were favorable, the system was found to lack support for non–native readers who attempt to read more challenging texts and would benefit if it improved how it explains and uses lexile scores. Over the trial period, comprehension on tests that immediately followed reading fell as text difficulty increased over time. Scores on the reading proficiency test provided by the system decreased slightly but significantly, throwing doubt on its validity over short periods of time. In addition to limitations of the system itself, reading improvement was found to be related to the level of texts which students chose to read, but not to how many or their length. The implication is that increased efforts must be made to convince students not to read at a level way below their personal lexile level when improving proficiency is the goal.

Keywords: Graded reading, lexiles, proficiency, fluency, extensive reading, EFL, TESOL

Article history: Received 31 July 2021, Revised 30 October 2021, Accepted 2 November 2021

1. Background

Reading is widely accepted as a crucial skill in foreign language learning [1], [12]; therefore, any resource, electronic or otherwise, that can promote it is to be welcomed. Furthermore, people learn to read, and improve other skills through reading, better if they independently read materials, typically out of class, that is graded to suit their current reading ability, either through selection or adaptation [1, 2]. Many studies have shown that this form of graded reading can improve reader comprehension [3, 4], as well as reading proficiency [5], writing skills [6], and vocabulary [3], [7, 8].

This study concerns the evaluation of a Lexiles–based online English reading management system, whose aim is to promote reading in just that way. It offers students a variety of graded texts to read, tracks which texts students read, and measures text comprehension and general proficiency.

1.1 Digital Reading Management and Lexiles

A complete system for achieving this, whether implemented by a publisher of graded readers as in the past [9] or by a digital reading management system

(DRMS) such as we consider here, requires three components, not all of which are fully understood or always successfully implemented [10]. It must measure (a) the difficulty of reading texts, and (b) the reading proficiency of readers, and it must (c) have an appropriate method of matching the two so that readers are offered texts at the most suitable level to help them improve. The widely advertised system of Lexiles promises exactly that [11].

Measuring text difficulty (a) has a long history of relying either on holistic expert intuition [12] or on selected objective indicators of difficulty as a whole, such as topic [13], vocabulary frequency [14–16], word and sentence length as used by ‘readability’ measures such as Flesch, Fog, and Coleman-Liau [17, 18], or more sophisticated measures such as the number of passives or clauses per sentence [19, 20], type–token ratio, and vocabulary frequency profile [21, 22]. Lexiles apparently rely just on word frequency and sentence length, omitting not only other linguistic measures but also relevant matters such as the topic, genre (e.g. linear narrative is usually easier), the number of pictures, etc. These measures then provide a useful but only approximate way of making or selecting progressively harder texts.

Reader proficiency (b) equally can be measured either by expert judgment, especially of a teacher rec-

*Corresponding author; email: schwo@hk.edu.tw

ommending a book to a pupil (e.g. [1]), or by objective tests [23]. The former has the advantage that factors like student knowledge of, and interest in, particular topics and genres can possibly be taken into account. It is still widely relied on, as evidenced for example by netTrekker d.i. [24]. A more objective approach to (b) involves actually testing reading proficiency of the student (whether an L1 or L2 speaker), e.g. by multiple choice reading comprehension questions after a short text is read, or from filling cloze gaps in a text. This side of the equation is less well developed [17], but a DRMS can only be effective and truly adaptive [25] if it incorporates regular reader proficiency testing, which a lexiles-based DRMS can.

The third and most controversial DRMS requirement (c) is automatising a connection between measures of reading proficiency and measures of text difficulty/readability. In the past, elaborate lists of claimed correspondences were used between the levels of text identified and scores on reading tests adopted by publishers [10]. A more recent approach exploits the increasing popularity worldwide of the Common European Framework of Reference (CEFR) as a reference scale with which language proficiency tests and exams of all types can be calibrated. It identifies and describes six levels, from beginner A1 through A2, B1, B2, C1 to C2 highly proficient. Websites such as *Text Analyser* [26] then use computerized text measures of word and sentence length, plus word frequency in the language, to assign reading texts also to CEFR levels. However, a system working with only six categories is not really refined enough to tell a learner precisely what text is just at their level for independent reading.

A lexile-based DRMS overcomes this by using a 0 to 2000 lexile scale both to assess reading text difficulty and to measure reader reading ability [27–29]. The precise system used to measure, and achieve parity between, lexiles as both a text and personal proficiency measure is in the hands of MetaMetrics Inc. The lexile system then comes with a recommendation for use thus. If a student takes a reading test score for example lexile 880L, then the system predicts that books with lexile scores from 100L below 880L to 50L above would be suitable. In that, it is claimed a reader should understand at least 75% of what they read. One such book, for example, is J K Rowling's *Harry Potter and the Sorcerer's Stone*.

This recommendation, however, must be seen as contentious. First, researchers suggest that comprehension scores of at least 92% [30], rather than 75%, are required to allow a text to be read without interruptions to look up unknown words and the like being so frequent as to destroy the reading process. In order to achieve that, a reader would in fact need to read a text with a lexile below their personal lexile by around 25%. For instance, a reader with lexile 880L would have to be matched with a book of 660L like *Arthur, for the Very First Time* by Patricia MacLachlan.

Second, the difficulty of the text relative to reader proficiency needs to be fine tuned in relation to the purpose of reading. Krashen [31] and Chiang [32] stated that acquisition occurs when the learner is exposed to comprehensible input, characterized as 'i+1', meaning input that is just a little beyond the learner's current proficiency level (i), so maybe +10L to +100L above learner calculated lexile level. This allows comprehension to occur but still offers some new language to be learned. For Krashen, input at i or i-1 (i.e. with a lexile below the reader's calculated lexile) will be comprehensible, but too easy and not enable learning to occur as a new language would not be introduced. On the other hand input at i+3 would not be understandable, so again not lead to acquisition. A similar concept is found in social theories of learning with the zone of proximal development, i.e., a zone just beyond the learner's current capability, in which social interaction may have a beneficial effect on acquisition,[33–35].

In contrast to the above, Nation [36], for example, makes the point that while input at i-1 or indeed i-3 may not help proficiency in the sense of language knowledge to develop, it can serve a useful separate role in developing fluency, here reading speed [37]. That is, material -100L or more below a reader's calculated lexile level, while not assisting language learning in the sense of increasing stored language knowledge, can be used to give students an easy read that allows them to become quicker at retrieval of what they know. There is additionally a third important kind of learner purpose, which is not to read for any language development purpose at all, but purely to enjoy or learn from the content of what is read. That is prominent for example in non-native students studying their major, such as medicine, nursing, or engineering, through the medium of English. Their ideal reading material would again be below their calculated lexile, but instead of increased fluency being the result, the desired outcome would be subject knowledge of their specialism.

Aside from those issues, limitations and shortcomings of lexiles have been voiced in several areas [38–41], [29]. These include criticism of the limited range of text indicator measures used, the cost of ascertaining the lexile of texts of teacher's choice, and the obscurity and predominantly L1 orientation of the personal lexile measurement. In a rare empirical study evaluating lexiles for non-native speakers, Holster et al. [40], for example, found that for Japanese learners, the Yomiyasusa system was a better predictor of English text difficulty than lexiles. However, that system seems to provide no systematic way of matching readers with texts-based on individual measures of texts and readers.

1.2 Objectives of the Study

In summary, lexiles are widely accessible, and ambitious claims are made for their usefulness. Criticisms have been made of them, but they are based largely on theory and argument: there is a paucity of evaluation of any lexiles-based DRMS through actual empirical trialling, especially with non-native speakers. For those reasons, the present study set out to evaluate a lexile-based DRMS as an aid to developing reading proficiency, through an empirical trial. The study, therefore, aims to answer two kinds of questions, concerning objective reading improvement and subjective perception of value by the students:

1. What patterns of development do we observe over the period of the study in terms of increased lexile of texts read, change in the amount of text read, and change in reading scores achieved?
2. What attitudes do the students have towards the lexile-based DRMS, especially in terms of its visual design, interface design, and usefulness for learning to read better?

These questions conform to a common view that the evaluation of teaching resources, including software, needs to consider both their objective and subjective merits.

2. Method

2.1 Participants

Forty-seven L1 Chinese sophomore English majors at Hungkuang university in central Taiwan agreed to participate in the reading intervention during the first half of a reading and discussion course; 44 completed the questionnaire. The group contained late beginners (CEFR A2), aged 19–23 (mean 19.7, SD 1.0), 70% female. Their mean lexile score was 716L (SD 171L) claimed as grade 3 reading ability in the US native speaker school system. They can be regarded as motivated students not least due to the fact that they have to pass the TOEIC test of English reading and listening proficiency at a 550 level in order to graduate.

2.2 The Reading Intervention Procedure

The DRMS vendor (Scholastic Taiwan) provided free use of its online system for three months. The students received a short orientation session in class from the vendor, explaining lexiles (both as text and personal scores). A teacher who had used the system also shared her students' experiences. Participants, next, took a pre-test of general reading proficiency in the DRMS to establish initial personal lexiles. That measured comprehension of short texts with multiple choice items in a progressively adaptive framework and calculated a personal lexile score. The class teacher then encouraged the students to make good use of the platform for reading their own choice of texts after class, either on or off-campus, to upgrade their reading proficiency. Neither the researchers nor the

DRMS limited the level of text offered in any way. Several classes were arranged in the lab with a TA to assist those who still had questions about the reading platform, but there was no researcher presence.

Shortly after the pre-test, it was found that most students were choosing a level of text to read that was well below their pre-test lexile. The instructor encouraged students to read books that matched their proficiency level or higher by giving extra credit. Furthermore, as an incentive, it was announced that the post-test at the end of the study intervention would replace the usual mid-term exam of the course.

Texts were presented by the DRMS as in hardcopy, without any associated activities that a teacher might supply in a reading class. While reading, the students sometimes had access to an audiobook version of the text to listen to but otherwise had no facilities on the site to support reading that presented linguistic or other challenges to comprehension.

The reading texts were authentic texts written initially for native speakers, including children, plus some "simplified" graded texts written for younger or elementary readers. Most of the texts students chose to read were expository/descriptive texts concerning culture-neutral universal general knowledge or interest topics such as *Galaxies*, often treated in a light way, e.g., *What am I?* and some with personal information value such as *Internet safety*. There were also a few fictional and biographic narrative texts. All texts were "general English", not academic English or English for specific purposes. This is, therefore, not the English of university subjects taught through the medium of English nor the English needed in specific jobs such as nursing which some students might later take up.

After reading each text, students were offered a quiz of up to seven multiple choice comprehension items. In order to assess progress in reading proficiency, a DRMS mid-test and post-test similar to the pre-test were given, respectively in the middle and at the end of the study intervention period. Shortly after the post-test (which was also the mid-term examination for the course) the questionnaire was administered to elicit an in-depth understanding of students' attitudes to the DRMS reading experience.

2.3 Measures

Aside from providing the post-text quizzes and the proficiency tests (validated by Metametrics), the DRMS also recorded the number of words read by each student. The software however credits a reader who reads even a small part of a text with having read it all. It also logs the number of texts read where the student completed the follow-up quiz, which can be seen as an indication of motivational intensity or interest, and the lexile levels of all texts (partially) read, from which we derived, mean lexile of texts read by an individual, and the difference between the lexile of

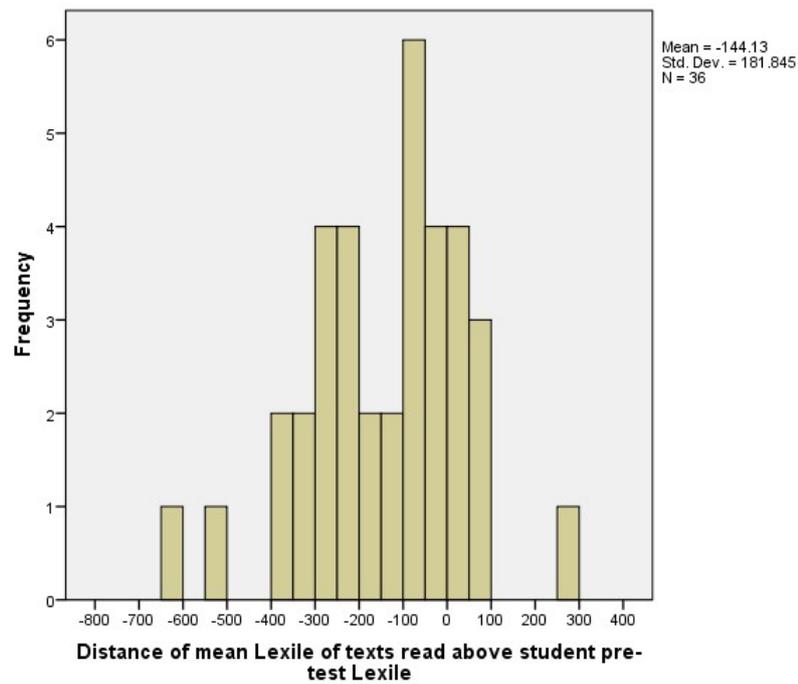


Figure 1: Histogram of differences between initial personal lexile scores and mean text lexiles.

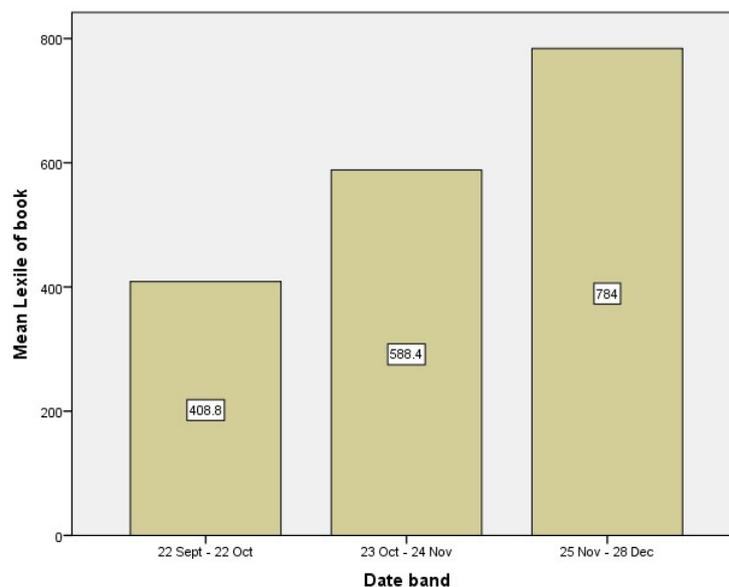


Figure 2: Lexile of book read over time.

a text read by a participant and their personal lexile level.

2.4 The Questionnaire

The questionnaire (modified from [42]), administered in Chinese, elicited demographic information and attitudes to features of the DRMS website, mostly in closed response mode with a five-point Likert scale. Those covered visual features of the DRMS website (QV, 15 items), website interface and organization (QI, 11 items), and learning benefits of the site (QL, 25

items). Closed items are listed in appendix 1. There were also three open response attitude items at the end.

2.5 Data Analysis

Since the data was not normally distributed, Spearman rho was calculated where correlations were needed.

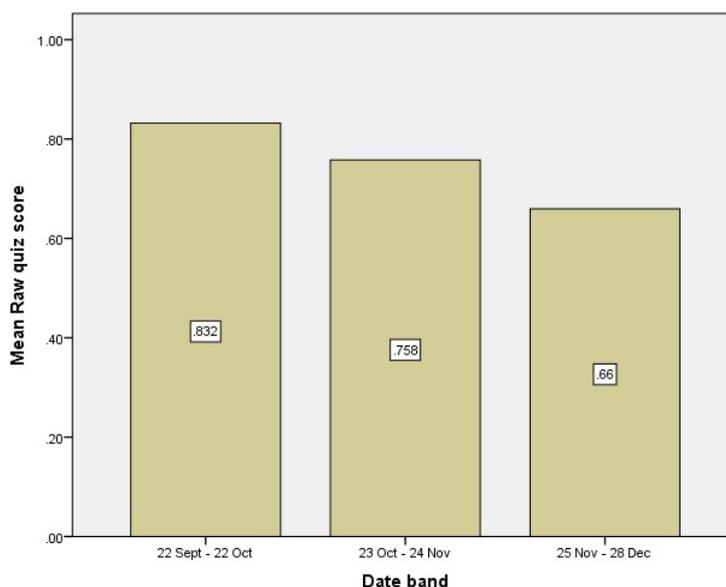


Figure 3: Raw quiz scores over time.

3. Results

3.1 General.

299 different texts were read by one or more participants. The most popular one was *I've Lost my Hat* (17 reads). However, less than half of the different texts (110) were read by more than one reader. This range testifies to the variety of choices offered, and used by students using the system. The 47 students read, on average, 12 texts each over the period of the study, ranging from 1 to 44 (SD 8.8).

The number of words in a text read by participants, as recorded by the software, ranged from 21 to over 85000, with a median of 700 words per text. As we have already mentioned, however, this may not be the number of words actually read. Lexile levels of texts read ranged from 10L (*Learn about Weather: Snow*) to 1170L (*Are we Alone?*), median 620L.

3.2 RQ1 Reading Development.

This can be revealed from changes in reading proficiency scores (personal lexiles), changes in immediate post-reading comprehension quiz scores, and from progression to reading more difficult texts.

3.2.1 Test scores of general English reading proficiency.

These scores, intended to measure learning, fell slightly over the period of the study, with means respectively pre-test 716L (SD 171L), mid-test 659L (SD 194L), and post-test 654L (SD 159L). This decline follows a significant downward trend ($F= 5.65$, $p= .022$), though with quite a small effect size (partial eta squared .127). It contrasts with the student perception in the questionnaire, which was that their reading had improved.

While this result was unexpected, there could be several reasons for it. One is that the time period of the study (three months) is quite short for one to expect a measurable improvement to occur: by contrast Vlangas [43] gathered data over two years to record an increase. Second, in the absence of reliability statistics for the DRMS test used, possibly the range of scores over time falls within the within-subject measurement error of the test. Third, in the absence to our knowledge of full publication by companies like MetaMetrics of how they determine lexile equivalences from proficiency tests such as those used by the DRMS, we cannot be certain of the validity of the personal lexiles given to the participants. Fourth, a lack of increase in reading proficiency, if not a fall, could be explicable if participants were reading texts below their personal lexile, possibly due to a wish to read for pleasure or fluency development rather than to push their boundaries and read to increase language knowledge. In a context where the DRMS did not constrain the texts offered to the reader, it emerged that this was happening to a great extent.

Figure 1 shows that students were reading texts considerably below their personal lexile level (mean=144L). Only four were reading on average in the official recommended area just above their calculated personal lexile, while the remainder were reading below, or in one case way above. We must, however, be aware that these figures are based on average lexiles for the texts read, so this does not mean that those reading on average at a lexile below their personal lexile did not, in fact, read some texts at an appropriate level along with others at a much lower level.

In the light of our discussion earlier, we may alternatively take the view that reading just above 75% of calculated personal lexile is appropriate for learning.

Table 1. The results for items in the section entitled Visual Design (QV).

	Mean	SD	t	p
1. The design and content of learning website are clear.	3.82	.815	6.66	<.001
2. The visual design of the learning website can help me to find the books that I want to read.	3.73	.899	5.37	<.001
3. The visual design of the learning website allows me to set my reading level.	3.95	.861	7.35	<.001
4. The visual design of the learning website offers different ways to find the books I want to read.	3.82	.922	5.89	<.001
5. The font size of the learning website is well suited to reading.	3.86	.979	5.86	<.001
6. The color of the learning website's background is subdued; it won't hinder and affect learning.	4.09	.802	9.03	<.001
7. The page setup of the learning website is consistent and unified.	4.00	.863	7.69	<.001
8. I think the overall visual color of the learning website is consistent with the design style.	3.95	.834	7.59	<.001
9. The visual design of the learning website is clearly visible.	4.02	.792	8.57	<.001
10. I like it that the learning website provides a playback function.	3.89	.813	7.23	<.001
11. The pictures in the learning website are related to the theme of the book.	3.84	.861	6.48	<.001
12. The visual effect of the learning website layout guides me in navigating the interface.	3.73	.845	5.71	<.001
13. The design of the learning website is too complex; I don't know how to use.	2.61	1.10	-2.32	.025
14. I like it that the learning website provides full screen function.	3.82	.870	6.24	<.001

Appendix 1. The questionnaire items and statistics

Scale 1= Strongly disagree to 5= Strongly agree. Significances are of difference between the observed mean and 3

On average that would be in a band such as -170 to -80L in Figure 1, where only 10 readers fall. In fact whatever band one chooses as the Goldilocks band for acquisition, at least two-thirds of our participants fall outside it reading too high or too low.

In order to see what was occurring at a more detailed level, we conducted a more detailed study of individual cases, of which we report two here which are particularly illuminating.

U104A128 obtained 529L (75% = 396L) on the pretest, but then read books in order with lexiles: 300, 490, 60, 270, 30, 110, 50, 10, 450, 540. Except for the very last, they were all below this student's lexile and many also below their 75% lexile so predicted to be easy reads, in some cases very easy, and indeed this student obtained 80–100% on most of the comprehension quizzes. This student dropped 46L points between the pre-test and post-test.

By contrast one of the bigger rises came from U104A143, who initially scored a lexile of 689L (75%= 517L), which rose by 87L in the post-test. This student read many texts (15), some of which were way below even 75% personal lexile. However, it is notable that six were close to or above the pre-test lexile (ranging 690–920L). Thus, the student received an appropriate challenge in the area where learning from input can occur, which maybe accounts for the satisfactory rise in personal lexile score.

Thus overall, the message seems to be that, for proficiency to increase (or in our study decrease less), the student needs to be challenged by texts at an appropriate lexile level close to and just above their lexile score

(or maybe their 75% lexile score), so as to achieve a learning benefit, consistent, for example, with the tenets of Mason and Krashen [5].

3.2.2 Text difficulty and Quiz scores measuring understanding of a text just read.

Figure 2 indicates that the lexile (i.e., difficulty) of the books being read goes up over time, as would be expected, even though student choices often remained below their initial calculated lexile. The average lexile of the books read in the second month was 180L higher than that of books read in the first month. In the third month, the average was 196L higher than in the second month, suggesting an increasing differential. Certainly, we would want any reading program to enable the students to read progressively more demanding texts over time.

The mean score for immediate post-reading quizzes (Figure 3, where 1 = 100% correct) was initially 83% (not far off the ideal 92%) but went down over time to a far less satisfactory 66%. This may be explained by the fact that the texts became harder. However, if reading ability was improving we would expect the comprehension scores for individual texts to remain the same or even improve. Therefore, this again indicates that underlying reading proficiency was not improving, so reading harder texts simply resulted in lower comprehension.

All these results are significant (taking as cases each text read by each person: the number ranges 477–552, depending on missing data). Correlation of lexile of

Table 2. The results for items in the section entitled Website Interface Usability (QI).

	Mean	SD	t	p
1. The learning site guide is well designed, allowing me to quickly find the function I want to use.	3.66	.914	4.79	<.001
2. The overall interface of the the learning website is well designed, so I don't have to spend a lot of time learning how to use the function.	3.57	.950	3.97	<.001
3. The functions of the learning site are very clear to see where I can click, and I can indicate the hyperlink.	3.75	.781	6.37	<.001
4. The learning website's menus, buttons and icons are easy to understand.	3.77	.859	5.97	<.001
5. After using the learning website, I don't have to explore how to use it.	3.82	.815	6.66	<.001
6. There are some functions of the learning website that I don't know.	3.02	1.11	.136	.893
7. The function icons in the learning website are unclear, and I often press the wrong link.	2.91	1.14	-.530	.599
8. The instructions on the the learning website are clear.	3.70	.954	4.90	<.001
9. The difficulty of the books provided by the learning website is clearly arranged.	3.73	.788	6.12	<.001
10. I think the various test modes on the learning website are easy to use.	3.91	.802	7.52	<.001
11. I think the function of the learning website is easy to understand.	3.66	.861	5.08	<.001

book with date was $\rho = .457$, $p < .001$. Correlation of quiz score with date was $\rho = -.209$, $p < .001$.

3.2.3 Correlations among reading-related measures.

Change in students' lexile between pre- and post-test was most strongly correlated with their initial lexile score ($\rho = -.366$, $p = .028$). This shows that students who entered the study with lower lexile scores (i.e. individual reading proficiency) improved their lexile more (or decreased it less) over the period of the study, compared with those who entered with relatively higher personal lexile (greater reading proficiency). This is heartening, although one would hope that a reading program would help readers of all degrees of proficiency to improve.

There was also a significant negative correlation between mean lexile of texts read and pre-post change in lexile score of a student ($\rho = -.332$, $p = .034$), showing that students who read easier texts also improved their lexile level more over the period of the study than those who read (on average) harder texts. This finding is related to the preceding one since initial student lexile scores were positively correlated with lexile of texts read, as one would hope ($\rho = .492$, $p = .002$). Those who started with lower reading ability read easier texts, on average, and those who started higher read harder texts, even if, as we saw earlier, the texts were still often below the calculated lexile levels of students at all levels (Table 4).

Change in student lexile between pre- and post-test was not significantly related to the number or length of texts read. This indicates that amount of reading is not crucial to proficiency improvement, but rather the lexile level of the texts read. Since, in the current study, the latter decision was left to the students, this may have contributed to the lack of strong positive per-

sonal lexile improvement scores over the period of the study.

3.3 RQ2 Student Attitudes.

Due to space limitations, only the key findings are considered here: for full results see appendix 1. All 15 attitude items in the section labelled Visual Design (QV) were significantly endorsed above the midpoint of the scale (3) (using the one-sample t-test). Particularly items 7 and 10 were rated above 4 (on the 1-5 scale), concerning the use of color and clarity of pictures, which accompany book covers and texts. In the set of 13 positively worded items about the interface design (QI) no items reached a mean rating of 4, but all were significantly positively endorsed above the midpoint. The two negatively worded items, concerning whether users got confused by the DRMS and whether the site lacks clear signposts, were not significantly agreed or disagreed with.

In the third main attitude section of the questionnaire, concerning pedagogical usefulness (QL), all 25 items were positively worded and significantly agreed with, though only one topped a mean of 4. This was item 24 which stated that the site allows students to read books on many different topics as much as they like. This testifies again to the range of choices offered.

Overall, the answers to the closed questionnaire items favor the website. However, students were not asked about crucial specific aspects such as the lexiles: whether they understood them, how they used them, and whether they saw them as useful. Nor were they asked about their experience moving from easy to harder texts: whether there was enough support to help them understand. Some of these issues were, however, raised in answers to the open questions.

Table 3. The results for items in the section entitled Learning Satisfaction (QL).

	Mean	SD	t	p
1. The learning website is helpful for me to learn English.	3.86	.765	7.49	<.001
2. The functions of the learning website meet my needs for learning English.	3.77	.859	5.97	<.001
3. Using the learning website, you can improve your English learning ability and use it more handily in the future workplace.	3.79	.833	6.23	<.001
4. I can strengthen my English learning through the learning website and get reasonable scores in the course-related assessment.	3.74	.848	5.76	<.001
5. I think using the learning website to study is helpful for me to study cross-culturally.	3.67	.892	4.96	<.001
6. I will share with others the content of the courses I have followed on the learning website.	3.67	.969	4.56	<.001
7. Learning through the website has made me more interested in English learning, and feel satisfied and a sense of accomplishment.	3.70	.887	5.16	<.001
8. If I need to improve my "reading" ability in the future, I will use the learning website as a practice tool.	3.81	.880	6.07	<.001
9. If I need to improve my "vocabulary" skills in the future, I will use the learning website as a practice tool.	3.65	.897	4.76	<.001
10. If I need to improve my "listening" ability in the future, I will use the learning website as a practice tool.	3.42	1.03	2.67	.011
11. The audio file in the learning website is helpful for my oral training.	3.47	.855	3.57	.001
12. The audio file in the learning website is helpful for my listening.	3.67	.892	4.96	<.001
13. The audio file in the learning website is helpful for my reading comprehension.	3.77	.868	5.80	<.001
14. The learning website provides a platform for me to take the initiative to learn English. It is an indispensable supplementary textbook after class.	3.67	.837	5.28	<.001
15. Through the learning website, I can enhance my learning ability.	3.84	.814	6.74	<.001
16. Through the learning website, I can enhance my reading comprehension skills.	3.93	.799	7.64	<.001
17. Through the learning website, the course learning method can enhance my motivation for active learning.	3.63	.952	4.33	<.001
18. Through the learning website, I can apply the professional knowledge I have learned to develop my potential in the course learning method.	3.74	.848	5.76	<.001
19. Through the learning website, I am learning English and I can integrate English into my life more.	3.84	.898	6.11	<.001
20. I am rich in the reading materials provided on the learning website.	3.79	.833	6.23	<.001
21. I am interested in the reading materials provided on the learning website.	3.81	.880	6.07	<.001
22. I am satisfied with the reading materials provided on the learning website.	3.84	.861	6.48	<.001
23. I think the advantages of the learning website outweigh the disadvantages.	3.86	.852	6.73	<.001
24. The learning website allows me to read books on different topics unlimited times.	4.02	.792	8.57	<.001
25. The learning website allows me to choose suitable texts according to my reading ability.	3.95	.806	7.86	<.001

Asked what improvements they would recommend, 45% said nothing needed improvement, with praise for the wide choice of reading, the value of the tests and quizzes, and clarity of the site. The remainder offered a wide range of suggestions. Some students called for more mainstream, interesting, or funny books to be included. Others targeted the text lexiles. In particular, there was a call for the lexile score to appear in the book list so it would be easier to pick the right level book. Similarly, another stated that the search engine did not make it easy to search books by lexile score.

Another useful and straightforward feature requested was a function allowing unfinished books to be saved in the library since at present it took a long time to locate any unfinished books.

While reading, some wanted more help with words, such as an English dictionary or words that were clickable to obtain help, maybe in Chinese. One wanted an increase in oral training. Concerning the tests, one person called for the lexile level test of the student to be available for them to take whenever they wanted. There was also a call for the quizzes to have the cor-

Table 4. Correlations among reading related measures.

		Pre-Post student lexile change	stu- dent lexile	Mean lexile of texts read	Number of texts read	Mean words in texts read	Mean of texts initial student lexile
Initial student lexile	rho	-.366*		.492**	.300	.052	-.546**
	p	.028		.002	.076	.762	.001
	n	36		36	36	36	36
Pre-Post student lexile change	rho			-.332*	-.264	.089	.039
	p			.034	.095	.582	.823
	n			41	41	41	36
Mean lexile of texts read	rho				.195	.533**	.404*
	p				.221	.000	.015
	n				41	41	36
Number of texts read	rho					-.175	-.243
	p					.275	.153
	n					41	36
Mean words in texts read	rho						.380*
	p						.022
	n						36

rect answers and explanations given.

4. Discussion and Conclusion

The aim of the study was to evaluate the lexile-based DRMS in an EFL context. Taking the students' subjective perspective first, overall, the DRMS was favoured, especially for its ease of use, color, and pictures, and the wide choice of reading offered. Possibly, however, the positive response was in part because they were mainly using the resource in a way that did not fully challenge them, by often reading below their personal lexile level, i.e., apparently reading for pleasure rather than for proficiency improvement. Nevertheless, the general clarity of signposting of links and functions could be improved and fuller listening resources provided. For our focus on learning to improve reading, however, the following two key issues emerged which, as far as we know, have not been reported before from any empirical EFL study of lexiles.

First, insufficient support is offered when a lack of comprehension occurs. Given that this DRMS is primarily used to provide educational resources for the native speaker market, this must not be overlooked when extending the learner audience to an EFL/EIL market, especially if the aim is to boost learner proficiency, which, according to the literature, means students must read texts a little above their personal lexile or comfort level (Krashen [31]). In support of independent reading, a computer resource has to supply the scaffolding and other support that otherwise a teacher would supply [44]. Relevant to the present case, the system should incorporate clickable words and phrases to offer help on word meaning, complex grammar, or unfamiliar names, terms, and concepts. Also, the system vendors should provide links that would allow students to access suitable reference

sources without leaving the reading page, such as a sidebar of dictionaries, including dictionaries of language and culture, and good quality bilingual dictionaries, rather than to Google Translate, which some students were observed accessing independently of the website during the trial. Arguably this lack of support as readers moved to harder texts may have contributed to the lack of proficiency development.

A second key area mentioned concerns the lexiles, which need to be more prominent at all points where lists of books are displayed or searched. It seems that in fact, some information on text lexiles is available that the students in this study did not manage to access due to unfamiliarity with some function keys, and this needs to be made more transparent. Furthermore, the student's personal lexile level should not only be displayed when they take a proficiency test, but the user should be reminded of this regularly and of the need to use their own lexile to guide what they read. This is critical if readers are to make better choices of what to read concerning their own personal lexile scores, which there was considerable evidence was not occurring properly in the study.

The objective measures of reading development using the DRMS did not yield unambiguous evidence that progress was really being made. Reading proficiency scores provided by the system tended to fall slightly, which puts some doubt on the validity of the tests used by the site to measure small changes in reading proficiency/personal lexile level, at least in the short term. Indeed, other studies that record an increase in personal lexile score over time generally took place over a far longer period (e.g. two years in Vlan-gas [43]). This is in fact consistent with findings of other studies of extensive reading that show that proficiency gain is slow [36]. A relevant factor also seemed to be the extent to which readers made good choices of

what to read in order to improve proficiency, i.e., texts just a little above their personal lexile level (or 75% of it). That is to say that they seem to have often chosen texts that were relatively easy for them, more perhaps for the pure enjoyment of the content without undue effort than to improve reading proficiency. This policy might, in fact, have coincidentally helped them boost reading speed/fluency, but that was not something we measured.

One conclusion from the above is that there are some clear limitations of the lexile-based DRMS, in the area of user support for example, which really need addressing if it is to be used successfully by non-native learners. However, the last mentioned issue that is so important in a way falls between the responsibility of students, teachers, and software such as the DRMS. This is the fundamental question about the real role of independent reading assisted by a DRMS. Arguably the students quite reasonably saw the purpose as different from that which the teacher had in mind: they saw it as a chance to use what English they knew to read for pleasure, without undue linguistic demands, with focus on content, while the teachers, mindful of the student low proficiency level, saw it as an activity aimed to increase student reading proficiency, so with some focus on language, rather like intensive classroom reading: for this reason paradoxically student subjective ratings were favourable although objective reading proficiency scores did not increase. However, the researchers agree with Krashen [31] and Nation [36] that these are not entirely compatible goals.

One way to deal with this is for the teachers to be more directive in guiding the students in these matters, as they tried to be in our study. However, ideally it is the learner who would make an informed decision autonomously about which kind of reading it is best for them to do at any given time. In that case, the teacher would assume the role of a trainer or advisor of the student rather than a dictator of what they do. In any event, the DRMS can then help by making sure that the choice of text does fit the aim of the student. As Bower [25] remarks, computer systems should be designed to support development at the level of the learner's abilities (their "zone of proximal development" or ZPD). Our study was not refined enough to say for certain whether that ZPD (or $i+1$ in Krashen's [31] terms) starts at the threshold of their personal lexile level or at something closer to 75% of their lexile level. That requires further research. However, once that is decided, we suggest a way forward, in either case, is for the site to configure what it offers more clearly into different distinct options where the software automatically constrains what texts can be accessed by the reader, relative to whatever is decided as the threshold personal lexile. From the reading research perspective, one might suggest three user-chosen modes.

Content mode would offer the reader an unconstrained choice of texts, and no time pressure, from a lexile range that they should be able to cope with (relative to their threshold personal lexile) and so allow focus on the content, whether read for pleasure or a necessary topic of their subject course. An attractive feature would be a chance for the reader to rate each text for interest (or "Like") after reading it, which could be used by the system to suggest what to read next. Fluency mode by contrast would have the goal of increasing reading speed (Nation [36]). It would offer texts in a range 100L+ below the student's threshold personal lexile score. However, when the student reads a text, some form of time pressure would be applied to prompt the reader to read a little bit faster than their initial speed. Thirdly, proficiency mode would only offer text choices a little above the reader's threshold lexile level, from 10L up to 50L above. There would be no time pressure, but the reader would meet unknown words and maybe longer sentences than they were used to and need to work to understand. In this mode, support features to combat a lack of comprehension would be important. There are in fact signs that the promoters of lexiles are moving somewhat in this direction with respect to L1 reading instruction. As reported by Vlangas [43], the lexile bands identified as corresponding to Common Core State Standards for reading at each grade in US schools have recently been extended upwards in what is called "stretch bands". Effectively this means more children will have to read texts that are above their personal lexile level, so reading proficiency should increase (Mason and Krashen [5]).

As indicated at the start, the whole system of lexiles itself raises many issues and has been attacked by influential critics (Krashen [38]; Hiebert [29]; Holster, et al. [40]; Cunningham, et al. [41]). In particular, the lack of consistent correspondence between lexiles and CEFR levels is worrying [45]. Our study has shown some specific further limitations of an existing lexile-based DRMS in relation to potential use by non-native speakers. However, we cannot condemn the whole system as useless. As we discussed, it can be improved and there is a need for teachers to play a role in training students in making good decisions about what kinds of texts to read, so that their reading meets their current needs. A good DRMS can then supply them with choices of texts that match their requirements. Finally, there is also a need for more studies in this area, conducted over a longer period than ours, and ideally with more participants.

References

- [1] E-Y. Jeon, R. R. Day, The effectiveness of ER on reading proficiency: A meta-analysis, *Reading in a Foreign Language* 28 (2) (2016) 246–265.
- [2] D. Tan Ai Lin, A. Pandian, P. Jaganathan, Encouraging ESL/EFL reading among lower proficiency students at the ter-

- tiary level: The use of graded readers, *Reading Matrix: An International Online Journal* 16 (2) (2016) 20–36.
- [3] L-F. Lin, The impact of text difficulty on EFL Learners' reading comprehension and vocabulary learning, *Journal of English Language Teaching* 2 (2) (2012) 14–24.
- [4] U. Wan-a-rom, The effects of control for ability level on EFL reading of graded readers, *English Language Teaching* 5 (1) (2012) 49–60.
- [5] B. Mason, S. Krashen, Extensive reading in English as a foreign language, *System* 25 (1) (1997) 91–102.
- [6] F. M. Hafiz, I. Tudor, Graded readers as an input medium in L2 learning, *System* 18 (1) (1990) 31–42.
- [7] J. Liu, J. Zhang, The effects of extensive reading on English vocabulary learning: A meta-analysis, *English Language Teaching* 11 (6) (2018) 1–15.
- [8] J. Poulshock, Extensive graded reading in the liberal arts and sciences, *Reading in a Foreign Language* 22 (2) (2010) 304–322.
- [9] G. Claridge, Graded readers: How the publishers make the grade, *Reading in a Foreign Language* 4 (1) (2012) 106–19.
- [10] V. Rodrigo, Graded readers: Validating reading levels across publishers, *Hispania* 99 (2016) 66–86.
- [11] MetaMetrics, Inc. <https://www.metametricsinc.com/Lexile-framework-reading>, 2020
- [12] S. A. Crossley, D. Allen, D. S. McNamara, Text simplification and comprehensible input: A case for an intuitive approach, *Language Teaching Research* 16 (1) (2012) 89–108.
- [13] A. Ellis, *A history of children's reading and literature*, Pergamon, UK, 1963.
- [14] M. West, *The New Method Readers for teaching English reading to foreign children*. Longmans, Green and Co, New York, 1930.
- [15] M. West, *A General Service List of English words*. Longman, Harlow, 1953.
- [16] E. L. Thorndike, I. Lorge, *The teacher's word book of 30,000 words*, Bureau of Publications, Teachers College, Columbia University, New York, 1944.
- [17] M. Fernández-García, Review of issues in computer-adaptive testing of reading proficiency, *Language Learning and Technology* 5 (2) (2001) 19–22.
- [18] S. A. Crossley, D. B. Allen, D. S. McNamara, Text readability and intuitive simplification: A comparison of readability formulas, *Reading in a Foreign Language* 23 (1) (2011) 84–101.
- [19] U. Wan-a-rom, Graded reading vocabulary-size tests for placement in extensive reading, *Reading Matrix: An International Online Journal* 10 (1) (2010) 15–44.
- [20] D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan, *Longman grammar of spoken and written English*, Longman, London, 1999.
- [21] P. Meara, H. Bell, P.Lex: A simple and effective way of describing the lexical characteristics of short texts, *Prospect* 16 (3) (2001) 5–19.
- [22] P. Meara, T. Fitzpatrick, Lex30: An improved method of assessing productive vocabulary in an L2, *System* 28 (2000) 19–30. doi:10.1016/S0346-251X(99)00058-5
- [23] Y. Sawaki, Comparability of conventional and computerized tests of reading in a second language, *Language Learning and Technology* 5 (2) (2001) 38–59.
- [24] L. Huff, Review of netTrekker d.i., *Language Learning and Technology* 12 (2) (2008) 17–25.
- [25] M. Bower, A framework for adaptive learning design in a web-conferencing environment, *Journal of Interac-*
- tive Media in Education* 2016 (1, 11) (2016) 1–21. doi.org/10.5334/jime.406,2016.
- [26] Text Analyser. <http://www.roadtogrammar.com/textanalysis>, 2020
- [27] D. R. Smith, A. J. Stenner, I. Horabin, M. Smith, *The lexile scale in theory and practice*, Final Report. MetaMetrics, 1989.
- [28] B. R. Blackburn, Best practices for using Lexiles, *Popular Measurement* 3 (1) (2000) 22–24.
- [29] E. H. Hiebert, Beyond single readability measures: Using multiple sources of information in establishing text complexity, *Journal of Education* 191 (2) (2017) 33–42.
- [30] B. Laufer, G. C. Ravenhorst-Kalovski, Lexical threshold revisited, *Reading in a Foreign Language* 22 (1) (2010) 15–30.
- [31] S. Krashen, *Second language acquisition and second language learning*, Pergamon Press, Oxford/New York, 1981.
- [32] M. Chiang, Effects of varying text difficulty levels on second language (L2) reading attitudes and reading comprehension, *Journal of Research in Reading* 39 (4) (2016) 448–468. doi.org/10.1111/1467-9817.12049
- [33] J. P. Lantolf (ed), *Sociocultural theory and second language learning*, Oxford University Press, Oxford, 2000
- [34] I. Bakhoda, K. Shabani, Bringing L2 learners' learning preferences in the mediating process through computerized dynamic assessment, *Computer Assisted Language Learning* 32 (3) (2019) 210-236. doi.10.1080/09588221.2018.1483950
- [35] X. Chen, D. Meurers, Linking text readability and learner proficiency using linguistic complexity feature vector distance, *Computer Assisted Language Learning* 32 (4) (2019) 418–447. doi.10.1080/09588221.2018.1527358
- [36] I. S. Nation, *Learning vocabulary in another language*, vb vc-Cambridge University Press, UK, 2001.
- [37] Y. Iwahori, Developing reading fluency: A study of extensive reading in EFL, *Reading in a Foreign Language* 20 (2008) 70–91.
- [38] S. Krashen, The Lexile framework: Unnecessary and potentially harmful. *California School Library Journal* 24 (2) (2001) 25–26.
- [39] S. White, J. Clement, Assessing the Lexile framework: Results of a panel meeting, Working Paper No. 2001-08, U. S. Department of Education Office of Educational Research and Improvement, 2001.
- [40] T. A. Holster, J. W. Lake, W. R. Pellowe, Measuring and predicting graded reader difficulty, *Reading in a Foreign Language* 29 (2) (2017) 218–244.
- [41] J. W. Cunningham, E. H. Hiebert, H.A.Mesmer, Investigating the validity of two widely used quantitative text tools. *Reading and Writing: An Interdisciplinary Journal* 31 (4) (2018) 813–833.
- [42] S. M. G. Chwo, Is Web 2.0 pedagogy sustainable and equitable? The example of Livemocha for teaching English as an International Language in Taiwan, *Studies in the Humanities and Social Sciences (HungKuang University)* 20 (2017) 275–300.
- [43] J. A. Vlangas, *The Stretch Lexile band: Potential implications for adolescents who struggle with reading*, Ph. D. Dissertation, Notre Dame of Maryland University, 2016.
- [44] J. D. Bransford, A. L. Brown, R. R. Cocking, *How people learn—brain, mind, experience, and school*, National Academy Press, Washington DC, 2000.
- [45] M. Smith, J. Turner, *The Common European Framework of Reference for Languages (CEFR) and The Lexile® framework for reading bringing more precision to language learning*, MetaMetrics, US, 2020.