A Comparison of Two Online Systems for Extensive Reading

PAUL COLLETT
Shimonoseki City University

This paper presents a comparison of two computerised assessment systems for extensive reading - Moodle Reader (MR), and XReading (XR). Given differences in how the two systems measure reading outcomes, to what extent are there differences in recorded outcomes for readers based on the assessment systems used? Drawing on data from system logs for three extensive reading classes at a Japanese university, two of which had worked with MR exclusively and one which had used MR and XR for one semester each, reading outcomes were analysed to see to what extent they varied for the different groups. Results showed similar patterns for quiz outcomes under XR and MR, but with better reading outcomes for the XR condition, where students read proportionally more words and fewer books, and failed fewer quizzes than in the MR condition. The differences observed in the data may be attributable to the different systems, but they are not of a scale that could be considered problematic. These results are discussed in light of learner goal orientations along with some of the issues that may be faced in using such assessment systems in an extensive reading course.

Collett, P. (2018). A Comparison of Two Online Systems for Extensive Reading. *Journal of Extensive Reading* 6(3). http://jalt-publications.org/content/ index.php/jer/

Keywords: computerized assessment systems, reading outcomes, learner goal orientations

Implementing an extensive reading (ER) program involves a number of challenges, as noted by Day and Bamford (1998). Practically, instructors and administrators need to ensure a range of suitable reading material is available and easily accessible, either in the learning space or library. Students must be able to access books often and easily if they are to read the amounts required for successful extensive reading. In addition, some system to check that reading is being completed may be necessary, especially if part of a required or for-credit course. All this needs to be done in a way that will be manageable for administrators and teachers often hard-pressed for time, and that will not be overwhelming for students. A number of studies suggest these types of challenges are often an obstacle for the implementation of ER courses (Huang, 2015, Macalister, 2008). One possibility for reducing complexity is to automate parts of the process, such as assessment of reading, which can be time-saving for both students and teachers. While there has traditionally been resis-

tance to the idea of assessing ER, more recent thinking is that there is in fact a place for assessment in the process (Day, 2015; Robb, 2015; Waring & McLean, 2015). An option for streamlining the assessment part of ER is Moodle Reader (MR), which can be used with the Moodle Learner Management System. A more recent alternative is Xreading (XR), which offers a complete online environment for ER.

Moodle Reader

Moodle Reader (http://moodlereader.org) is a plug-in module for the Moodle learning management system, developed as a way to assess reading progress of graded readers and other books aimed at language learners. To use, MR must be installed on a computer system running Moodle, and configured via this system, and is accessed through a networked web-based interface (note there is also a variant of MR, M-Reader, intended for those without a Moodle setup available to run MR, but who require the same functionality. This version will not be discussed here). Reading material for students is not provided via MR, rather it offers a way to assess reading progress through timed quizzes, which can be taken by students once they have finished reading books. The quizzes are designed to test whether the student has actually read the book or not; they are not designed to test comprehension or memory of fine details. The module incorporates quizzes on over 4500 graded readers and books for young readers, with questions drawn from a quiz bank so that each student receives a different set of randomised questions, usually 10 for each quiz. Additionally, there is a configurable time-limit for quiz completion. (Robb, 2015, Robb & Waring, 2012). Taken together, the MR developers claim these features are such that students are able to take the quizzes at a time and place that suits them, with minimised possibilities of cheating (Day, 2015; The Moodle Reader Module, n.d.).

XReading

XReading (http://xreading.com) is a relatively new (launched April 2014) web-based online service for Extensive Reading. It is designed to make ER both easier for students, and for teachers and course administrators. The XR system consists of an online virtual library (the XR VL) and a learning management system (LMS). Through the VL, users can access and read graded readers using the web browser on an Internet-enabled device such as a smartphone or tablet or personal computer. The LMS enables the instructor or course administrator to view student and class progress, set reading assignments, and control such things as the level and genre of books available in the VL. A succinct overview of features is provided by Milliner and Cote (2014).

"Reading is its Own Reward": Objections to Assessment?

Both of these systems focus on assessing reading, although XR can be used without quizzes whilst still allowing for the tracking of reading process. However, the idea of tracking and assessing reading progress may give some cause for concern here as it is in contradistinction to one of Day and Bamford's (1998, 2002) 10 guiding principles for ER; namely that "reading is its own reward," and reading in an ER program should not be assessed. This has, like the other nine principles, developed into something of a commandment (Macalister, 2015; Robb, 2015; Waring & McLean, 2015) which has made some unwilling to use a system such as MR, or its variant, MReader (Robb, 2015). However, the ex-

tent to which these ten principles should be followed has been open to considerable debate since their introduction, and recently we have seen moves to re-conceptualise the principles in light of both theoretical and practical considerations related to doing ER (Day, 2015; Waring & McLean, 2015). Additionally, considering the publication in 2004 of a volume of followup activities for ER edited by Bamford and Day, along with Day's (2015) support for MR, and it would seem that the principle of reading as its own reward is more of a variable component (Waring & McLean, 2015) of an ER course, or one of a number of guidelines (Macalister, 2015), rather than something that needs to be adhered to (Day, 2015). Requirements of running an ER course in formal educational contexts should also be considered. Assessment is often expected, students assume they will be rewarded for work they complete, and progress must be quantified and reported to students and administrators. In these circumstances, having an automated, relatively easy form of measuring ER progress becomes more than just a possible component of a program, but rather an essential tool to effectively run such a program.

Assessment, Motivation and Goals for Reading

One concern that arises when considering a system to assess reading relates to goal orientations of learners, and the potential effect of relying on quizzes or assessment for learners—what kind of achievement goal orientations will this encourage? Briefly, achievement goal orientations have been posited as existing within a 4-way framework of mastery-approach, performance-approach, mastery-avoidance and performance-avoidance (Kaplan & Maehr, 2007; Fryer & Elliot, 2008). Mastery-approach orientations lie behind intrinsic motivation, the learners have a personal interest in attaining proficiency in a field above and beyond any externally-imposed outcome expectations. Performance-approach orientations are seen in learners making an effort to reach a goal not because they have an intrinsic desire to, but because achieving the goal serves a more utilitarian purpose and once they achieve this it is unlikely they will devote much effort to furthering the initial goal pursuit. Avoidance orientations are seen in learners who feel incapable of achieving goals, and who will direct their energies towards shifting the locus of poor outcomes away from themselves so as to avoid personal responsibility for said outcomes. This latter type of orientation is often tied to learning contexts which overly emphasise normative outcomes or which focus on comparisons between learners in assessment. While mastery-approach goals are the type that should ideally be promoted in learning environments, performance-approach goals can lead to better learning outcomes, at least over the short term (Elliot & Moller, 2003; Senko, Hulleman, & Harackiewicz, 2011), and the optimum approach to helping learners develop positive goal orientations may be to adopt a multiple-goal approach. Here the instructor would work on promoting mastery goals as desirable but also recognising that normative comparisons are inevitable in formal learning environments, and that performance-approach goals can be beneficial for academic outcomes. One important point that arises from the research into the 4-way achievement goal framework is the need to maintain an appropriate balance, not over-emphasising the performance aspect over the mastery side (Fryer & Elliot, 2008). A more recent conceptualisation of achievement goals has expanded the model to a 3 x 2 framework (Elliot, Murayama, & Pekrun, 2011).

This model includes task-approach and self-approach components, corresponding to successful completion of tasks and personal development respectively; initial research in validating this model suggests task-approach goals may be most effective in promoting classroom-based learning. However, this model is still in the early stages of development, so for the purposes of this paper, any reference to goals will be based around the 4-way framework.

Mori (2015) addresses the related issue of extrinsic and intrinsic motivation to read, suggesting that many learners may lack intrinsic motivation, reading instead "for reasons that are external to both the activity of reading and the topic of a text" (Mori, 2015, p.130). Furthermore, in a culture like Japan, learners may prefer having assessment exercises for ER. The point here is that Day and Bamford's (1998, 2002) suggestions regarding reading for its own reward may be unrealistic in certain learning contexts, and that assessing reading outcomes may be more beneficial for learners in formal learning environments than an approach that stresses reading as its own reward. That said, while accepting that performance-approach goals are likely going to be the dominant type pursued by students, using a quiz-based system for learner feedback on reading progress should still be able to contribute to the development or validation of a mastery-approach orientation where used appropriately. This includes delivering feedback in ways that do not promote a classroom atmosphere wherein students are encouraged to try to better one another in their reading outcomes.

Research on Moodle Reader

While MR, and the related M-Reading system are in widespread use (Keith, 2016; McBride & Milliner, 2016) research into their use is somewhat limited. The overall consensus from the available research suggests the system can be a useful addition to an ER course (Day, 2015). Robb and Kano (2013) found that an ER program implemented at their tertiary-level institution in Japan using the MR quiz system was successful in raising language proficiency of students in comparison to those that had not done ER. They pointed out the value of MR in enabling the implementation of a large-scale ER program. Weatherford and Campbell (2016, 2014) carried out a number of studies to see how the students at their particular institution, a women's university in Japan, rated using M-Reader as an assessment method for ER compared to written book reviews. Their findings suggested that the assessment system students were first introduced to tended to be rated as most preferable, but that a preference for the M-Reader system was related to more enjoyment and perceived value amongst students for ER. Curtis (2015) in a small-scale survey of 29 Japanese university students regarding the use of M-Reader found that the majority of respondents enjoyed using the system, and did not find the quizzes to be too difficult. Bieri (2015) reported generally positive attitudes of Japanese university students in his year-long ER course towards using M-Reader's quizzes, and noted that it helped in overall management of the course. Positive results were also reported by McBride and Milliner (2016), with students at a Japanese university surveyed at the end of a year-long ER course using M-Reader agreeing that the system was easy to use (59%) and that the quizzes were understandable (72%).

Research on XReading

To date, as is the case with MR, there

seems to be little published research into the XR system, let alone anything comparing the outcomes of this with MR. Milliner and Cote (2015) reported on Japanese high school students' opinions of XR and MR after a semester-long course where two classes used one of the two systems exclusively. Students in the XR group (n=20) reported a generally more positive perception of ER than the MR group (n = 24). Additionally, the XR group members were all able to reach the course reading goal of 50,000 words, compared to only 33% of the MR group reaching their goal of 30,000 words. Cote and Milliner (2015) found that Japanese university students (n=95) who completed a fifteen-week ER program using XR had a generally positive attitude to the system at the end of the course. Milliner (2017) reported that a sample of Japanese university students (n=19) using XR were able to read relatively large quantities of words; in this case 17 of the subjects recorded word counts exceeding 250,000 words over the course of an academic year, higher reading outcomes than for studies with a corresponding focus where different systems had been used to assess reading outcomes.

It should be noted that these studies are not looking at groups who used both systems, but rather comparing users of one system to the other. The present study takes a different approach, comparing reading rates for a group of learners who used both MR and XR, as will be outlined in the following sections.

Context of the Study

The data collected for this study was collected from 1st-year English majors in a four-year undergraduate degree course at a university in Southwest Japan. All the students in the first year were required to take a year-long ER course, of which the author was the instructor. Students were coming into the class at the beginning of their first year at university with quite low reading speeds (Swanson & Collett, 2016), and informal comments and classroom reactions of students repeatedly suggested a general belief amongst them that reading was difficult and not very enjoyable. Consequently, the aim of the ER course was to help students to become faster, more fluent readers, as well as to build confidence in their reading abilities. In class this involved activities focused on increasing reading speed, and intensive reading activities to help learners develop effective reading strategies. There was also an outside-class reading (ER) component loosely following Day and Bamford's (1998, 2002) ten principles of ER, but closer to more recent conceptualisations of ER (Macalistair, 2015; Waring & McLean, 2015). This ER component was spread over 15-16 weeks, and required the students to complete weekly readings of books from the university library graded reader / young readers collection. Other than advising students to choose books that were at a suitable level for them, and explaining how to do this, no restrictions were placed on the titles students could read. Students were free to borrow whatever books they liked, and to read them in their own time. Based on recommendations from research (Waring, 2011; McLean, 2014), the main goal of the course was to try and read 250,000 words in the course of the semester. To achieve this, students were advised at the beginning of the course to read for 180 minutes a week outside class. While these are attainable goals with effective planning, in reality few students read for this long or reached the 250,000-word level.

One important concern in running the course was how to track the reading prog-

ress of this ER component. This was necessary for a number of reasons: to ensure the students were reading regularly and at a suitable level, and to keep a record of the number of books and words read. So as to make the assessment system as easy as possible for the students, and to allow them to freely read from a broad range of books, as well as to reduce course administration loads, the MR module was initially used as a way to record progress. Having completed a book, students were expected to log into their account on the school Moodle system, and then to test their knowledge of the book read via the MR module.

Rationale for XReading

After using MR for two years, a number of small but significant problems with its usage had become apparent. Predominant here was an ongoing issue with students reading books borrowed from the library that did not have corresponding quizzes in the module. In many of these cases the instructor had to accept student claims that they had read and understood a book, and manually assign them an approximate word count. While we chose to believe students were being honest when reporting these incidents, it is one area where they could attempt to artificially inflate their word scores if so inclined. Working with library staff to produce a list of the books which do not have quizzes available and removing these from the ER library would seem like an ideal way to alleviate this problem; unfortunately, time and other constraints did not allow for this to be done during the course of this study.

We also found that students were sometimes unable to find a quiz in the MR system due to errors when doing title searches, or problems with understanding what series the books were categorised under. Different versions of the same titles also caused confusion, with students sometimes taking a quiz that did not match the particular book they had read. While there were ways around most of these problems, they did require the instructor to be easily accessible outside of class to help students with quiz problems so they could do the quizzes soon after reading, and it invariably meant some amount of class time had to be spent on addressing quiz issues.

Other problems, unrelated to MR itself, included students sometimes forgetting to bring books to class for reading activities. There was also the issue of students failing to return books to the library on time. This usually resulted in restrictions on further borrowing, and in some extreme instances these students with overdue books were unwilling to go to the library at all, due to concerns about how the library staff would react to their lack of respect for the library rules.

XR, with its online book library and quiz system, seemed to offer a solution to these problems. Plans were also in place to trail a summer vacation reading project, and as many students would not have physical access to the university library over the summer break this would have been something difficult to implement without a solution like XR. More generally, there was an interest in seeing how students would respond to XR, and if it would streamline the management of the ER component of the course. Thus, it was decided to trial the system.

Having made the decision to trial XR, implementation was set for the last month of the academic semester, July 2015. By this stage, students were familiar with the extensive reading component of the course

and were reading at a good rate. This timeframe also allowed students approximately a month before the end of the first semester to familiarise themselves with the system and to start moving over from physical print copies of books to those in the virtual library, something that would be necessary over the summer break when they would be required to do a holiday reading assignment.

The initial reaction to the shift to the new system was generally positive, with students interested in the possibility of reading on their smartphones. However, there was some initial discontent about the usage fee for XR - at the time, a one-off charge of approximately 1,300 yen for seven months – but after some class discussion about the functionality of the system, students agreed it seemed to be a worthwhile expense. Some students did express initial doubts about using an online system for reading, citing concerns about possible negative effects on their eyes or the difficulty of using a computer-based system; it was made clear to these students that they could continue to read print books and use MR for their quizzes if they preferred. It was also made clear to students that the XR system was accessible using a web browser on any kind of modern internet-accessible device, and that they were not expected to access it solely on their smartphones. This said, the majority of the students seemed to prefer using smartphones over other devices, and based on informal observations, it seemed had been using their smartphones to do the MR quizzes. It should be note however, that limitations in the logging functions for MR did not allow a clear way to verify this.

As the course instructor, there was some concern about the types of outcomes students would attain using the two systems. Given the differences in the quiz styles, would one system allow students to record higher word counts than the other? Would reading on electronic devices be somehow more difficult than reading physical books, leading to decreases in reading rates? To see if the change in systems led to large disparities in reading outcomes, it was decided to carry out a comparative analysis of the reading results for the two systems.

Method

Data were collected from the system reading logs of MR for 2013, 2014 and 2015, and XR for 2015. The 2013 (N=44) and 2014 (N=40) classes serve as a control group against which the outcomes for the 2015 class (N=31) can be compared. Participants in the classes were English majors at an all-women's 4-year college in Japan. The students were all in their first year of study, aged between 18 to 19 years old except for one mature student in the 2013 year group. The reading logs were downloaded in CSV format from the respective system servers at the end of the 2015 academic year. Slightly different information was captured in the logs depending on the reading system used, but general information such as number of books read by each student, word count per book, quiz score, and reading level was consistent enough across systems to allow for comparisons. Cases where the entries for a book were incomplete or where word counts had been manually assigned were removed, as were cases where a book had been read outside the start and end dates of the course terms. The data were analysed using the R language and environment for statistical computing (R Core Team, 2016).

It should be pointed out that with our configuration, a pass for a book, which is as-

sumed to show the book was read and understood, is a 60% score in the book quiz. This is based on the recommended settings for the MR system which are intended to allow students to pass the quizzes with a reasonable level of comprehension (Truscott, 2017). However, it is important to note that the quizzes for MR have 10 questions (q) selected randomly from a pool of questions where pool$q \geq 10$, while the XR quizzes are comprised of five questions selected from a pool where pool$q = 5$. This raises the concern that the quizzes in XR may be easier to pass than in MR. Given this concern, the aim of the research is to see how the two assessment systems compare. Specifically, the objective is to see if there are noticeable differences in the patterns of reading and the rate of passed quizzes (books read) for the XR group compared to the MR groups.

Results

Three different measures will be considered: word totals, words read per book, and number of books read. It is hoped this will make clear the patterns of reading for each group, and exemplify salient differences where they may exist. Inferential statistics are presented where appropriate, but the focus in the results is on graphical representations of data and confidence intervals (CIs) for the means. CIs are regarded as an alternative to statistical significance tests to show differences between means, and recommended for use in publications (Cumming, 2014; Larson-Hall & Plonsky, 2015; Plonsky, 2015). Similarly, graphical plots are advocated for their explanatory power (Hudson, 2015).

Rate of adoption

Before comparing reading patterns for the MR and XR conditions, it would be helpful to see if there is enough data to warrant a comparison. Figure 1 illustrates the rate of adoption of XR compared to the MR system. Initially, usage of XR was somewhat limited, with less than half of the 32 students using the system in the first two months following its introduction. Students were also expected to do some reading during the summer holiday period

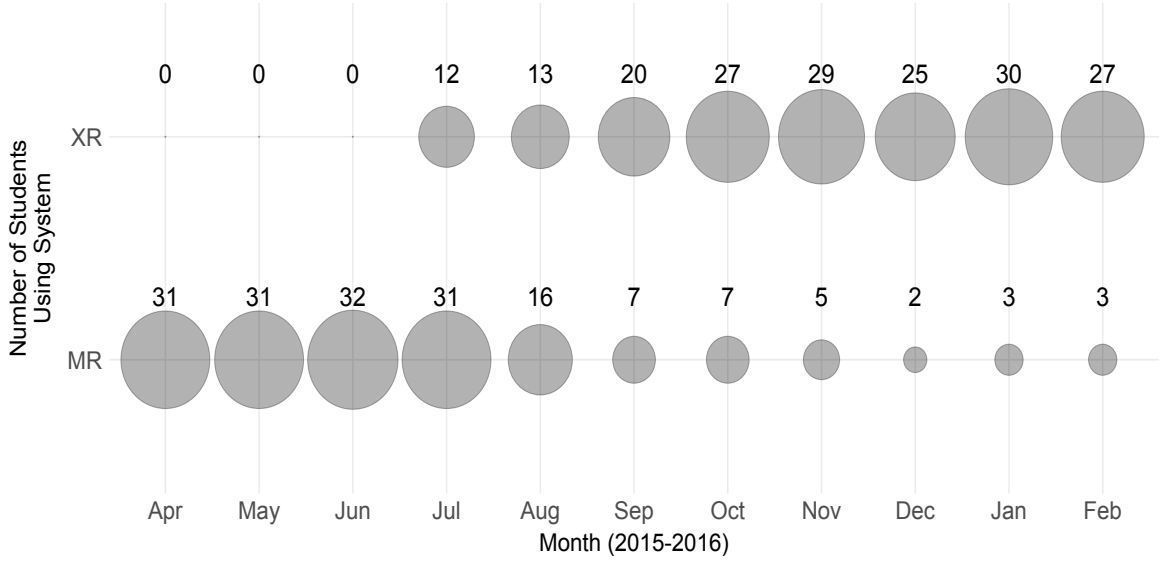

Figure 1: Comparison of usage rates for MR and XR

from August to late September; as many of them did not have access to the university library during that time, this accounted for an increase in usage of XR to nearly two thirds. Usage then increased once the second semester began, with a small number continuing to use MR - either in tandem with XR or exclusively. A period of two weeks in December when there were access problems with the XR server may account for the slight decline in usage in December, but this could also be explained by the end of year holiday period.

By the end of the 2015 academic year there was a 97% adoption rate with all but one student using XR (the single holdout failed to do any reading using either system), although a number of students were also working with MR and library books. It should be noted that in some months some students did no reading, although they had used either MR or XR prior to this, meaning that at the end of the second semester, all students who had completed the ER component of the course had used XR, with the majority using this exclusively. To what extent did the outcomes recorded under XR compare to those of MR? This will be covered next.

Quiz Totals

A comparison was carried out of the counts of total quizzes passed and failed for 2013, 2014 and 2015 classes. Table 1 shows the proportions of quizzes passed and failed, along with the mean words read per book, for first-year students using MR in the first and second semesters for 2013, 2014, and MR and XR in the first and second semesters of 2015 respectively. Figure 2 presents box plots of the total number of passed and failed quizzes for each year group. Each year group has a decrease in the number of quizzes taken in the 2nd semester, with increases in proportions passed. The difference in proportions between the two semesters is smaller than that for the other two years, at approximately 5% as compared to approximately 18% and 10% of data for 2013 and 2014. While the box

Table 1
Quiz Outcome Proportions and Means Words Read Per Book

Year	Term	N	Total Quizzes	Proportion Quizzes Passed	Proportion Quizzes Failed
2013	1	44	2,233	74.56	25.44
	2	44	1,679	88.21	11.79
2014	1	40	1,152	75.26	24.74
	2	40	925	82.92	17.08
2015	1	31	1,206	81.26	18.74
	2	31	892	84.98	15.02

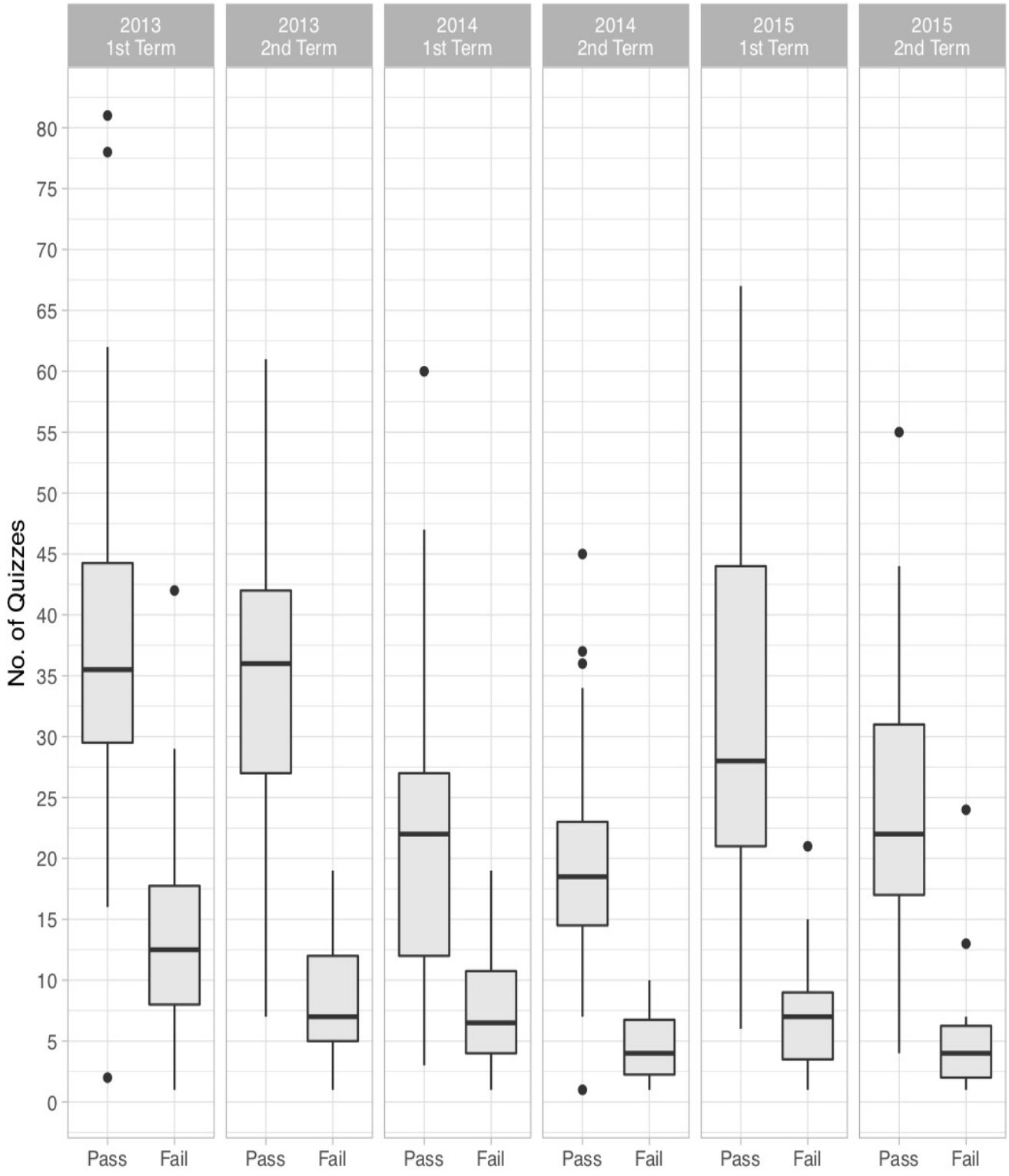

Figure 2 Box plots of quizzes passed and failed by class. 2015 Term 2 is the XR group.

plots show similar patterns in quizzes passed and failed for each year, there were more quizzes passed by the 2015 group in the second semester when using XReading than for the previous year groups. However, given that the 1st semester total for 2015 is also larger than the previous years, the data points to overall similar rates of outcomes for each year for the number of quizzes passed and failed.

Words read Per Book

Figure 3 consists of violin plots of the distribution of the word length of each book read by students for each term. In a violin plot, the width of the plot indicates the distribution of the data (i.e. a bulge in the plot means a larger number of measures are clustered around that point), with the length indicating the complete range of responses. In this case, the graphs have been truncated to remove outliers to better display the key data. The median is indicated by the black circle, the mean by the white diamond. Means, medians and 95% CIs are given in Table 2. Immediately noticeable is the change in the reading patterns for the second term of 2015 - overall, students are passing more quizzes for higher level books, with a mean increase of 32.7% compared to 16.3% and 4.8% for 2014 and 2013 respectively. Note also the decrease in the mean and median of the failed quiz data for 2015, which goes against the previous trends.

A mixed ANOVA was used to test for the within and between effects of the term and year group on the reading total data. Field and Wilcox (2017) recommend that due to problems with outcomes of classical ANOVA tests on non-normal data, robust versions of statistical significant tests should always be carried out unless there is clear evidence of the normality of the data set. With the outliers present in this data, it can be assumed it does not meet conditions of normality; this was confirmed by inspection of Q-Q plots, as well as a Shapiro-Wilk test of normality for which all groups of data where sta-

Table 2
Mean and Median Words Per Quiz

Year	Term	N	Mean Words Read: Passed	Mean Words Read: Failed	Median Words Read: Passed	Median Words Read: Failed
2013	1	44	2,013	3,912	1,120	1,700
	2	44	2,109	3,683	1,291	1,700
2014	1	40	1,956	3,698	1,109	1,400
	2	40	2,277	3,924	1,228	1,785
2015	1	31	1,863	2,581	1,224	1,275
	2	31	2,473	2,347	1,694	1,340

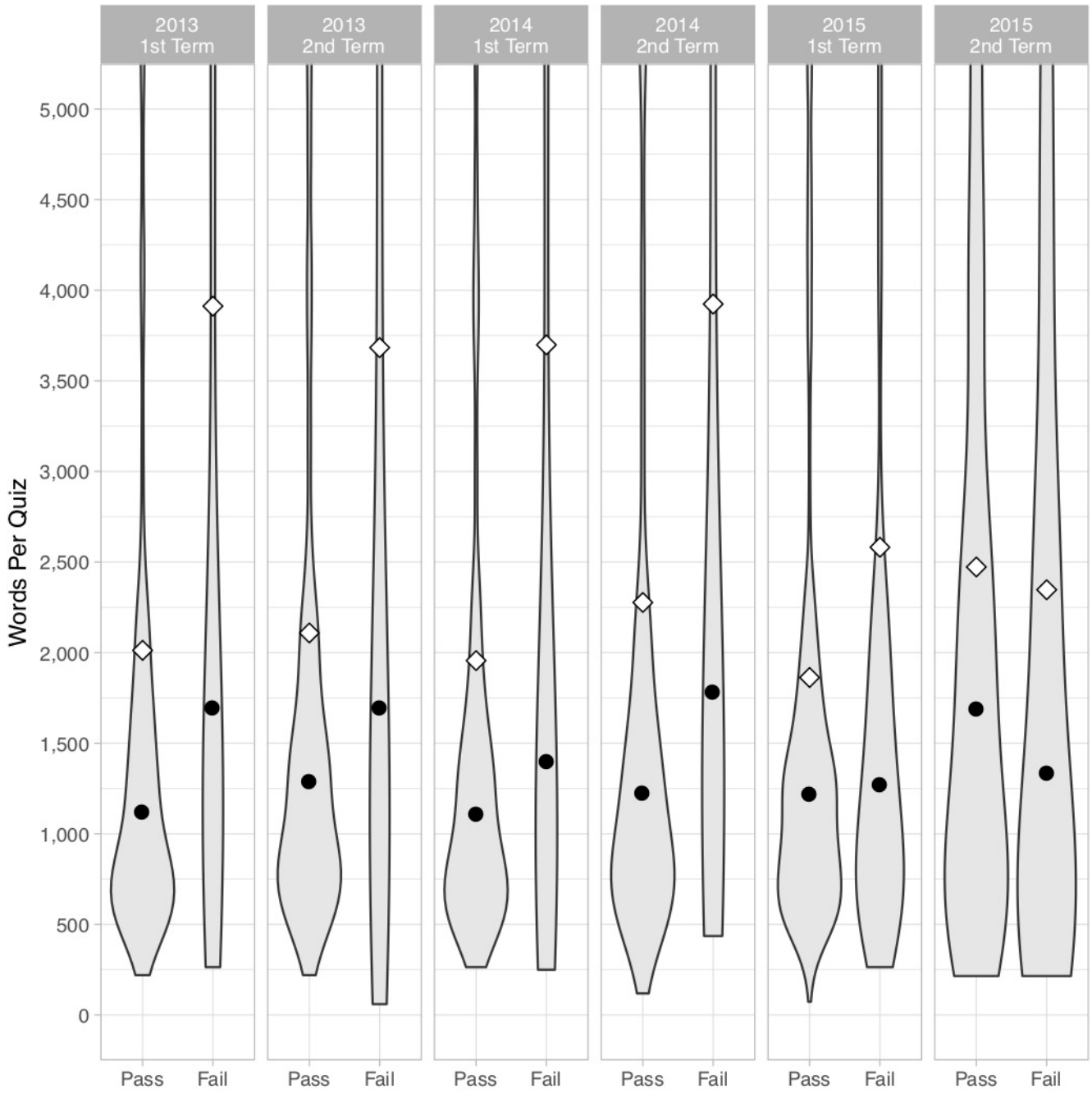

Figure 3: Distribution of mean word length of books read per term. 2015 Term 2 is the XR group

tistically significant at $p < .01$. Due to this violation of the assumption of normality, a robust ANOVA was performed on the passed and failed data separately, along with tests for between-subject, within subject, and interaction effects. The tests were carried out using the bwtrim function of the WSR2 package in the R statistical package (R Core Team, 2016). For the passed data, the main ANOVA (20% trimmed means) gave a statistically significant result for term (18.21, $p < .01$) but neither for year nor the interaction of year and term. The functions sppba, sppbb, and sppbi were also calculated; these give the main effects and interaction effects of the factors using an M-estimator and bootstrap. The main effects were non-statistically significant at the $p < .05$ level, however the interaction effect of year by term was statistically significant ($p = .036$). In the case of the mean failed data, the ANOVA gave a statistically significant result for term (7.35, $p < .01$) and the interaction of year and term (9.1396, $p < .01$).

Once again, the main effects were non-statistically significant at the $p < .05$ level, but the interaction effect of year by term showed statistical significance ($p < .01$). The statistically significant result for the passed data can likely be accounted for by the large change in the mean number of words per quiz for the 2015 data as compared to the other two years. The results for the failed data can likewise be accounted for by the much lower outcomes for the 2015 year group than those seen in the 2013 and 2014 classes.

Total words read

Figure 4 shows violin plots of the total number of words read for each year. The mean and median are indicated in the inset (white) box plot by the diamond and horizontal bar respectively. Outliers who had considerably higher read reading totals than the other students have been removed; there was one such case for each year group, each of these students had read amounts approaching 200,000 words or more. Descriptive statistics for cumulative totals are given in Table 3.

Table 3
Cumulative Words Read as Measured by Passed Quizzes

Year	Term	N	Total	Mean	10% Trimmed Mean	Median	SD	Min	Max
2013	1	43	3,144,426	73,126	69,384	56,284	45,038	2,897	175,325
	2	43	2,919,027	67,884	67,198	62,652	30,451	8,535	160,347
2014	1	39	1,519,568	38,963	32,134	25,905	37,528	2,781	176,002
	2	39	1,547,088	39,669	33,995	29,076	34,496	602	173,409
2015	1	30	1,552,857	51,762	48,105	44,880	33,677	5,511	142,960
	2	30	1,701,142	56,705	53,490	47,351	35,342	5,718	165,628

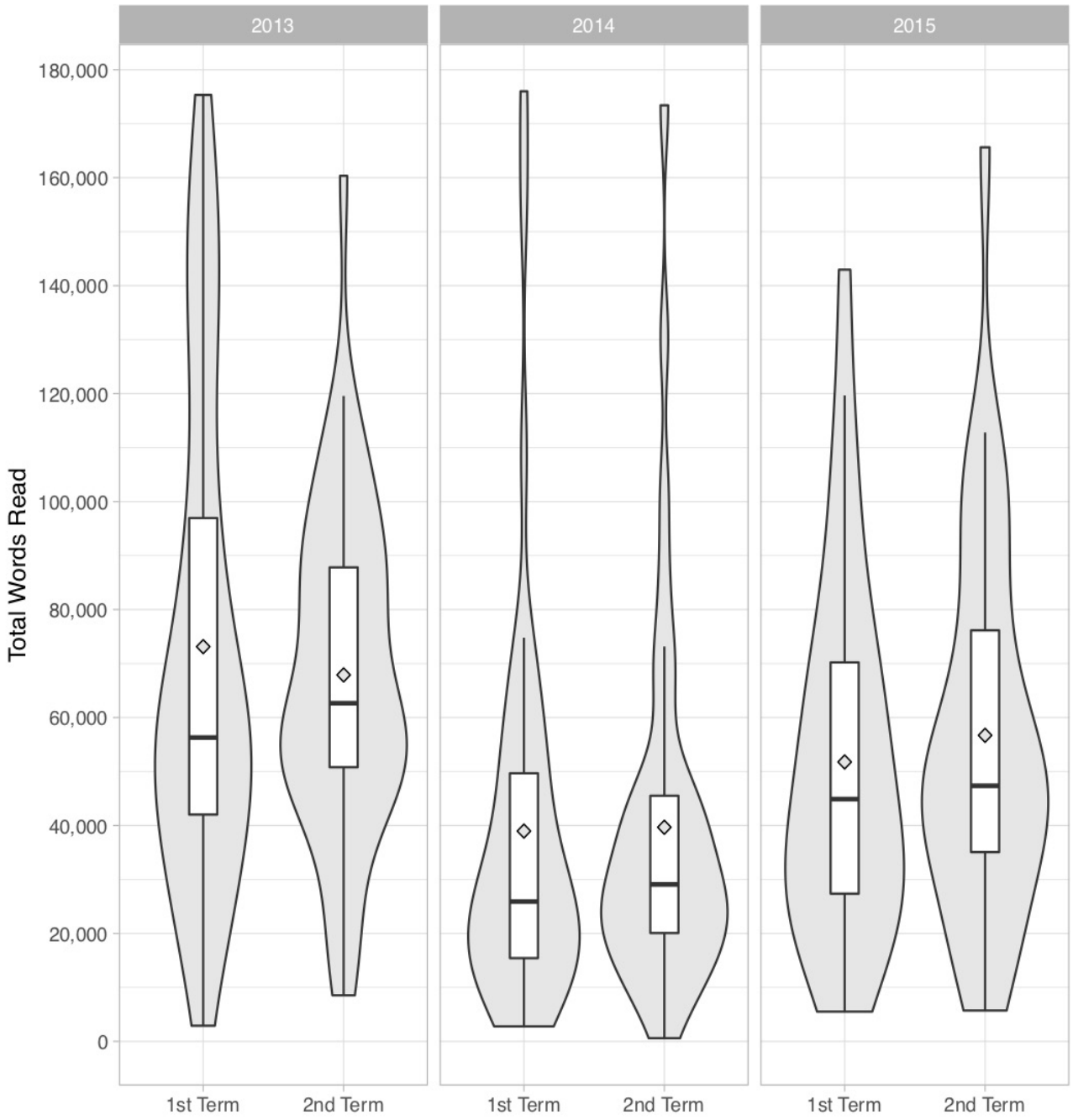

Figure 4: Distribution of total words read per term

Overall, the data distributions follow similar trends over the three years, with increases in the median number of words read in the second term for each group. Looking at the 10% trimmed means suggests that these changes are not due to skew from outliers. Although the trends are similar, there was a much larger proportional change in the total number of words read between the first and second terms of 2015; a 9.5% increase compared to a 7% decrease for 2013, and a 2% increase for 2014.

As an added check, an ANOVA was applied to the 2nd term data for each year. A visual inspection of the data with a Q-Q plot showed the data displayed non-normality confirmed by Shapiro-Wilk tests where $p < .01$ for all groups. Based on this, a robust between-measures ANOVA on 20% trimmed means bootstrapped 1000 times was carried out, using t1way from the WRS2 package in R. This gave a statistically significant result, $F(2,40.16) = 18.9$, $p < .01$, with an explanatory measure of effect size, $\xi = 0.5$, a large effect (Wilcox & Tian, 2011). Results of post hoc tests on the 20% trimmed means (using the function lincon) are given in table 4; as figure 4 suggests, we can reject the null hypothesis of no difference between the means for the 2013–2014 and 2014–2015 comparisons. The 2013 and 2015 comparison is inconclusive – as the CIs cross zero, it suggests there is no statistically significant difference between the data, although we have $p < .05$. However, the point to note here is that the distribution of the total words read in the second term of 2015 does not show any extreme differences in comparison to the other years.

This said, there was a considerably higher increase in total words read between the two terms for 2015 compared to the other years. Figure 5 shows the means and CIs for this group adjusted to account for shared variance of repeated measures (Field, Miles & Field, 2012; Morey, 2008). As the confidence intervals cross, it can be concluded that the within-year differences for the total number of words read in the first and second terms in 2015 are statistically indeterminate.

Overall, then, the results for the total number of words read present a pattern of the 2015 group showing greater improvement in their reading than the previous two years, with proportionally larger increases in the 2nd semester. As with the other measures of reading, the pattern of the data distribution is similar to that of the previous year groups, yet the results do not allow us to unequivocally say that the outcomes are due to anything other than chance.

Table 4
Post Hoc Test Results for Total Number of Words Read

	psihat	ci.lower	ci.upper	p.value
2013 vs. 2014	36510.15	21886.68	51133.63	< .001
2013 vs. 2015	15050.24	-3295.44	33395.916	.048
2014 vs. 2015	-21459.91	-39419.35	-3500.466	.005

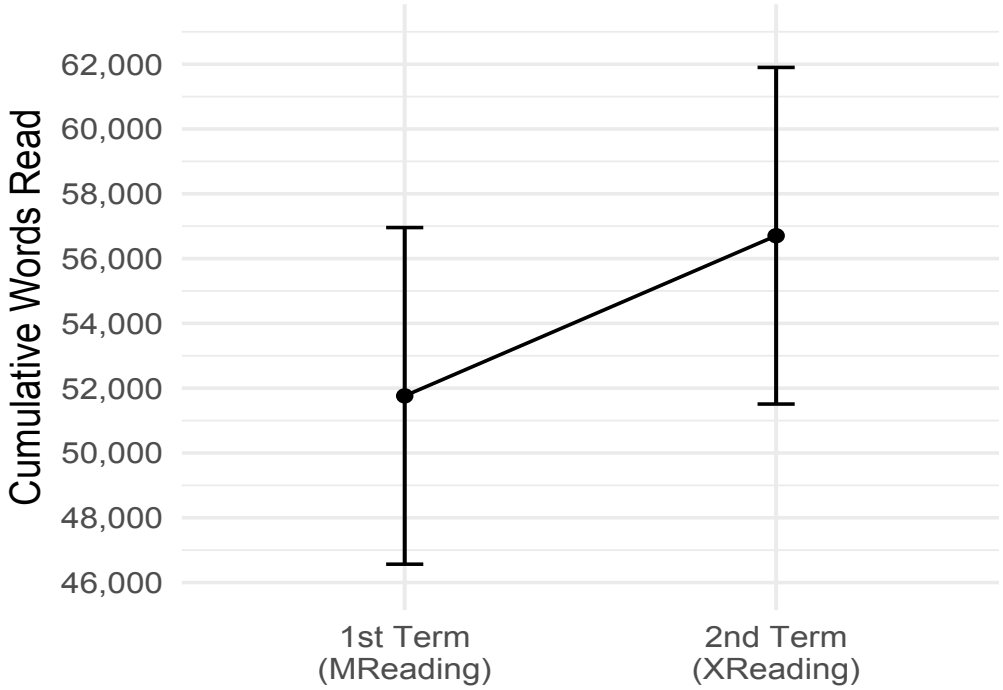

Figure 5: Means and 95% confidence intervals for the total number of words read in 2015

XReading Reading Speeds

One potentially useful function of the XR system is the ability to measure reading speed. This allows the instructor to see cases where reading speeds may be problematic - either extremely slow, or overly high. Approximately 5% of the results for the total 2015 XR were at reading speeds of 300 wpm or more; a reading speed higher than that of the average native speaker (Grabe, 2009). Of these cases, 56% passed the book quizzes, so approximately 3% of the total reading scores were for cases in which students were reading at more than 300 words per minute (the average reading speed here was 999 wpm). As these results were at odds with the reading rates observed in the classroom, they were not counted towards the final reading outcomes reported in this study. A similar, more extreme result was observed in the 2016 academic year, where the same group of students continued to use XR in their second year. For the single semester course in which they did extensive reading, 22% of the students recorded reading speeds of 300 wpm or higher. Of this number, 70% passed the quizzes. In other words, approximately 15% of all passed results were at reading speeds greater than 300 wpm, with an average speed here of 1,152 wpm. We will return to these outcomes in the following discussion.

Discussion

The aim of this study was to see how the XR and MR systems compare in terms of assessment of reading outcomes. Is there reason to believe that the two systems will provide similar results? Based on the analysis of the reading logs, the general trend

from the three year groups is consistent - the number of passed quizzes decreased in the 2nd semester for each year, with a median increase in words read. Changing to the online-only XR system led to increases in reading outcomes which were higher than in previous years. Students read proportionally more words and fewer books than in the MR condition. It may be that the changes observed are a result of learner development through extensive reading, a consequence of improvement in students' reading ability over the course of the ER program.

Other observations showed there was a smaller decrease in the number of quizzes failed for the 2015 group, along with a decrease in the mean number of words in failed quizzes, an opposite trend to the results from the previous two years. One possibility for this is that students in the 2015 year group were choosing to read easier books than the students in the earlier groups, and in the process passing more quizzes and increasing their word count, rather than attempting more difficult books with a higher possibility of failure. Carrying out an analysis of the levels of books read would be a logical step here, and a useful line of inquiry for further study. Overall however, without further investigation we cannot discount the possibility that the shorter quizzes used in the XR system may have led to the higher reading outcomes than those recorded using the MR system for assessment.

One important consideration arises regarding the high reading speeds recorded with XR. Are the students really understanding the books they are reading, or merely being able to demonstrate understanding easier through XR? Similar results for reading speeds were observed in a later course, and when questioned about this, these users reported they were reading books borrowed from the library but using XR for assessment as the quizzes were shorter and perceived as easier to pass than the MR quizzes. Obviously whether or not this accounted for the 2015 results is speculation, but it may be a reason behind the findings. Other reasons for the results could be that the system was not recording reading speeds correctly, or that students were rapidly flicking through a book to access the quiz without having read it, and did not necessarily need knowledge of a book's content to be able to pass the assessment that purportedly tests for knowledge of this content. If students are reading books from the library and then doing the XR quizzes, it somewhat defeats the purpose of using XR. Alternatively, if they do not need to complete the books to answer the quizzes, we then need to question the veracity of all the reading results. A recent account of ways in which students work to maximize assessment outcomes with minimal reading (Tagane, Naganuma, & Dougherty, 2018) highlights the problematic nature of self-assessment systems, and in tandem with some of the findings here, it brings us back to the original question guiding this study—should reading be assessed, and how?

A key component of the course was providing students with weekly feedback on their reading progress and their outcomes, and it may be that overly focusing on this, and by association, the course assessment is negatively contributing to learner goals which could lead to an over-emphasis on performance and the avoidance of failure. This is something that should be addressed in future studies, as the current data does not provide a means to assess learner goal orientations. Another issue to consider is that the extensive reading course requires

the students to plan and structure their reading on their own, and it may be that this level of autonomy is not suitable for some learners who, rather than devoting time to reading, will find ways of appearing to have read instead. The problematic reading speed data reported in the results suggests this may have been the path some of the 2015 year group students chose to follow, effectively choosing performance avoidance goals to achieve high reading outcomes. In light of Tagane, Naganuma, and Dougherty (2018) this is something that needs to be controlled for in a course where students are given responsibility for their own learning.

Problems Encountered

The biggest issue encountered in using XR was the instability of the system. Between late October and early December there were a number of incidents ranging from hours through to days when it was inaccessible. To their credit, the system administrators were usually quick to address these problems, but these outages still left times when it was unclear what had happened to the system and when it would be back online. Not having direct control of the administration of the system, or information on the status of technical problems, and the need to rely on the administrators to provide clear details about the scale of problems, as well as be able to get the system functioning again, proved to be quite frustrating. Unfortunately, these are limitations of any technological innovation, often out of the user's control. One of the possible consequences of system instability, especially if users are new to the system, could be user discontinuation (Rogers, 2003), and I had expected to see a percentage of students switching back to library books and MR following these problems. That this did not happen either says a lot for the strengths of XR, or possibly more likely, that the students were happy to have a break from the reading assignments.

It is hoped that the various server issues and outages that limited student access will not be a reoccurring feature. In all likelihood, students of university age and below have grown up used to "always-on" access to information. XR promises this, and in this particular case study, mostly delivered. However, if there are problems it could cause users to focus on these to the detriment of the advantages of the system, as for them the advantages are aspects they take for granted. Students assume that a system like this will work and always be available. It has been endorsed by the course instructor, and assumedly, on an institutional level, and it is something they have paid for the right of access to. And here is perhaps an issue that may make or break the application, striking a balance between getting enough subscribers to cover infrastructure costs to deliver a smooth user experience while at the same time being able to scale up to meet any unexpected demands. For these reasons, while recognising the usefulness of XR, one should also be mindful of potential pitfalls. It would be advisable to have a backup system in place, not just in the eventuality of access problems, but also to counter the possibility that not all students will be in a position to do the reading fully online. Additionally, it needs to be stressed that committing students' reading data to a system over which the instructor does not have full control could lead to problems accessing that data when required. Ensuring reading logs are saved (and backed up) regularly is of course one way to stay prepared for any unexpected surprises here.

Another potential problem is the lack of

books in the XR virtual library. With only around 470 titles at the time of the study, it did mean that books at certain levels are limited in availability. For the students in this study, who worked with the system over about 6 months, this was not a serious issue, but it could impact on courses extending over a number of years.

Limitations of the Study

One of the biggest limitations is the difference in the number of questions of the quizzes students read, as well as a lack of an indication of the reliability and validity of the quizzes for both the systems. Whilst data on pass and fail rates for individual titles could be compared across students, without a standardized reference point for the students' English level it would be difficult to draw any strong conclusions. For this reason, trying to assess reliability or validity has not been attempted in this study.

Another limitation is that the study does not include any measure of fluency by which to compare the reading gains to see if the outcomes were actually contributing to improvements in the learners' ability to understand what they were reading. Of course, the concern of the study is not with increases in fluency per se, but it would strengthen the findings if this had been addressed.

Having data from only one semester for the XR condition is another weakness. Completing a follow-up study with a wider dataset to ascertain if the results found here are replicable across different groups should provide richer data on which to base further conclusions.

Conclusion

To date, little research has been published into the use of XR and its role in assessing ER. It is hoped that this paper demonstrates that XR can contribute to a reading program with similar outcomes to those seen in using the MR system for assessment.

XR is still a relatively new system, and its features are somewhat limited. Many are being added, but as with the adoption of any system, check that it meets your needs before committing.

What can be said is that using XR to deliver content and track progress for an extensive reading program, especially if the reading component is to be carried out outside of class, was a relatively successful approach. Whether or not the choice of the course instructor is to assess reading, XR provides a convenient platform to allow students to read, and for both students and instructors to keep track of their outcomes.

For the instructor, it solves one issue some may face if using the other main online ER assessment system, MR (or MReader) - that of a mismatch whereby books available in the school library may not have corresponding quizzes available. And XR extends the convenience offered by MR by moving the whole reading process online, potentially eliminating the need for a large physical library of readers—although the number of titles in the virtual library at time of press is perhaps still too low to run a long-term reading program.

Based on the results of this study, assessing an extensive reading course with XR will likely give slightly better outcomes than if using the MR quiz system. Whether students need to read an entire book to pass the quizzes in both systems is something that may be an issue. Having

alternative ways to check this would be necessary if there are doubts that the assessment outcomes are a true reflection of student progress, however this may add unwanted complexity to a reading course. Structuring and running the course to ensure the goals of the course do not excessively promote the assessment component while prioritizing the other "commandments" of ER is one possibility here.

Of course, the usual cautions stand about the extensibility of the findings. Further research in other learning contexts will help provide a better picture of the benefits and drawbacks of the system. Additional research into how assessment systems such as XR influence learning goal orientations would also be useful to better understand how assessment of extensive reading can best be carried out.

References

Bamford, J., & Day, R. R. (Eds.). (2004). *Extensive Reading Activities for Teaching Language*. Cambridge: Cambridge University Press.

Bieri, T. E. (2015). Implementing M-Reader: Reflections and reactions. *Extensive Reading in Japan, 8*(2), 4-7.

Cote, T., and Milliner, B. (2015). Implementing and managing online extensive reading: Student performance and perceptions. *IALLT Journal of Language Learning Technologies, 45*(1), 70-90. http://ialltjournal.org/index.php/ialltjournal/article/view/67/58

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science 25,* 7–29. http://journals.sagepub.com/doi/abs/10.1177/0956797613504966

Curtis, T. (2015). MReader quizzes: Help or hindrance? *Studies in the Humanities and Sciences, 51*, 75-82. http://id.nii.ac.jp/1080/00002284/

Day, R. R. (2015). Extending extensive reading. *Reading in a Foreign Language, 27*, 294-301. http://nflrc.hawaii.edu/rfl/October2015/discussions/day.pdf

Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language, 14*, 136-141. http://nflrc.hawaii.edu/rfl/October2002/day/day.html

Day, R. R., & Bamford, J. (1998). *Extensive Reading in the Second Language Classroom*. Cambridge: Cambridge University Press.

Elliot, A. J., & Moller, A. C. (2003). Performance-approach goals: good or bad forms of regulation? *International Journal of Educational Research, 39*, 339-356. http://www.sciencedirect.com/science/article/pii/S0883035504000345

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 x 2 achievement goal model. *Journal of Educational Psychology, 103*, 632-648. http://psycnet.apa.org/record/2011-13620-001

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage

Field, A. R., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy, 98*, 19-38. http://www.sciencedirect.com/science/article/pii/S0005796717301067

Fryer, J. W., & Elliot, A. J. (2008). Self-regulation of achievement goal pursuit. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and Self-Regulated Learning: Theory, Research and Applications* (pp. 53-75). New York: Routledge.

Grabe, W. (2009). *Reading in a Second Language: Moving From Theory to Practice*. New York: Cambridge University Press.

Hudson, T. (2015). Presenting quantitative data visually. In Plonsky, L (Ed.) *Advancing Quantitative Methods in Second Language Research* (pp. 78-105). New York: Routledge.

Huang, Y. (2015). Why don't they do it? A study on the implementation of extensive reading in Taiwan. *Cogent Education, 2*, 1-13. https://doi.org/10.1080/2331186X.2015.1099187

Jeon, E. Y., & Day, R. R. (2015). The effectiveness of core ER principles. *Reading in a Foreign Language, 27*, 302-307. http://nflrc.hawaii.edu/rfl/October2015/discussions/jeon.pdf

Larson-Hall, J., & Plonsky, L. (2015) Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning 65* (Suppl. 1), 127–159. https://doi.org/10.1111/lang.12115

Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review, 19*, 141-184. http://www.jstor.org/stable/23363939

Keith, B. E. (2016). Making quizzes for M-Reader and the MoodleReader quiz module. In M. Gobert (Ed.), *Proceedings of the 3rd World Congress on Extensive Reading* (pp 79-86). http://erfoundation.org/ERWC3-Proceedings.pdf

Macalister, J. (2015). Guidelines or commandments? Reconsidering core principles in extensive reading. *Reading in a Foreign Language, 27*, 122-128. http://nflrc.hawaii.edu/rfl/April2015/discussion/macalister.pdf

Macalister, J. (2008). Integrating extensive reading into an English for Academic Purposes program. *The Reading Matrix, 8*, 23-34. http://www.readingmatrix.com/articles/macalister/article.pdf

Mair, P., Schoenbrodt, F., and Wilcox, R. (2016). *WRS2: Wilcox Robust Estimation and Testing*. R package version 0.9-2. https://cran.r-project.org/web/packages/WRS2/

Maxwell, J. A. (2005). *Qualitative Research Design: An Interactive Approach*. Thousand Oaks, CA: Sage Publications.

McBride, P. & Milliner, B. (2016). Introduction to M-Reader: An online extensive reading aid for schools. *The English Teacher, 45,* 96-105. https://journals.melta.org.my/index.php/tet/article/download/80/16

McLean, S. (2014). Evaluation of the cognitive and affective advantages of the Foundations Reading Library series. *Journal of Extensive Reading, 2*, 1-12. http://jalt-publications.org/content/index.php/jer/article/view/5/5

Milliner, B. (2017). One year of extensive reading on smartphones: A

report. *The JALT CALL Journal, 13*(1), 49-58. http://journal.jaltcall.org/articles/13_1_Milliner.pdf

Milliner, B., & Cote, T. (2014). Effective extensive reading management with Xreading. *The Language Teacher, 38*(6), 32-34. http://www.jalt-publications.org/tlt/departments/tlt-wired/articles/4184-effective-extensive-reading-management-xreading

Milliner, B., & Cote, T. (2015). Comparing two approaches to extensive reading (ER) management in the Tamagawa ELF program: M-Reader and Xreading. 教師教育リサーチセン ター年報 第 [kyoushi kyouiku risaachi sentaa nenpou], 5, 115-123.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4,* 61-64. http://pcl.missouri.edu/node/63

Mori, S. (2015). If you build it they will come: From a "Field of Dreams" to a more realistic view of extensive reading in an EFL Context. *Reading in a Foreign Language, 27*, 129-135. http://nflrc.hawaii.edu/rfl/April2015/discussion/mori.pdf

Plonsky, L. (2015). Statistical power, p values, descriptive statistics, and effect sizes: A "back to basics" approach to advancing quantitative methods in L2 research. In Plonsky, L (Ed.) *Advancing Quantitative Methods in Second Language Research* (pp. 23-45). New York: Routledge.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rogers, E. M. (2003). *Diffusion of Innovations* (5th Ed.). New York: Free Press.

Robb, T. (2015). Quizzes—A sin against the sixth commandment? In defense of MReader. *Reading in a Foreign Language, 27*, 146-151. http://nflrc.hawaii.edu/rfl/April2015/discussion/robb.pdf

Robb, T., & Kano, M. (2013). Effective extensive reading outside the classroom: A large-scale experiment. *Reading in a Foreign Language, 25,* 234-247. http://nflrc.hawaii.edu/rfl/October2013/articles/robb.pdf

Robb, T., & Waring, R. (2012). Announcing MoodleReader version 2. *Extensive Reading World Congress Proceedings, 1*, 168-171. http://erfoundation.org/proceedings/erwc1-Robb-Waring.pdf

Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist, 46*, 26-47. http://www.tandfonline.com/doi/abs/10.1080/00461520.2011.538646

Swanson, M., & Collett, P. (2016). Researching EFL Learner Reading Speed Gains. *Bulletin of Seinan Jo Gakuin University, 20*, 99-116.

Tagane, Y., Naganuma, N., & Dougherty, P. (2018). Academic dishonesty in extensive reading programs: Stories and strategies from student interviews. *The Language Teacher, 42*, 9-12.

The Moodle Reader Module. (n. d.). The Moodle Reader module. http://moodlereader.net/moodle/

Truscott, G. (2017). A closer look at the MReader settings. *Kinki University Center for Liberal Arts and Foreign Language Education Journal, 8*, 17-32.

Waring, R. (2011). Extensive reading in English teaching. In Widodo, H. & A. Cirocki (Eds.) *Innovation and Creativity in ELT Methodology*. Nova Publishers: New York. http://www.robwaring.org/papers/waring_Nova_2011.pdf

Waring, R., & McLean, S. (2015). Exploration of the core and variable dimensions of extensive reading research and pedagogy. *Reading in a Foreign Language, 27*, 160-167. http://nflrc.hawaii.edu/rfl/April2015/discussion/waring.pdf

Weatherford, Y., & Campbell, J. (2016). An Evaluation of Progress Measurement Options for ER Programs. In M. Gobert (Ed.), *Proceedings of the 3rd World Congress on Extensive Reading* (pp 55-68). http://erfoundation.org/ERWC3-Proceedings.pdf

Weatherford, Y., & Campbell, J. (2015). Student assessment preferences in an ER program. In P. Clements, A. Krause, & H. Brown (Eds.), *JALT 2014 Conference Proceedings* (pp. 661-668). Tokyo: JALT. https://jalt-publications.org/files/pdf-article/jalt-2014proc_071.pdf

Wilcox, R. & Tian, S. T. (2011). Measuring effect size: a robust heteroscedastic approach for two or more groups. *Journal of Applied Statistics, 38,* 1359-1368. http://www.tandfonline.com/doi/abs/10.1080/02664763.2010.498507