Reading in a Foreign Language October 2013, Volume 25, No. 2
ISSN 1539-0578 pp. 234–247

Effective extensive reading outside the classroom: A large-scale experiment

Thomas Robb and Makimi Kano
Kyoto Sangyo University
Japan

Abstract

We report on a large-scale implementation of extensive reading (ER) in a university setting in Japan where all students were required to read outside class time as part of their course requirement. A pre/posttest comparison between the 2009 cohort of students who read outside of class and the 2008 cohort who did no outside reading shows that the implementation of ER resulted in highly significant gains. A plug-in module for Moodle called "MoodleReader" was used to hold the students accountable for their reading. A new distinction between *replacement* ER and *additive* ER is introduced.

Keywords: Moodle, assessment, homework, outside reading, extensive reading

There have been many studies over recent years concerned with the effectiveness of extensive reading (henceforth ER), but virtually all of these studies are concerned with how ER works as an alternative to other activities that can be conducted in the classroom. They are concerned with optimizing classroom time to make learning more effective. In this article, we will review the most relevant literature concerning studies in English as a foreign language (EFL) environments, and then discuss an issue that has not yet been addressed—the effectiveness of ER conducted outside the classroom. Finally, we will introduce the MoodleReader program and how it was used to implement ER at Kyoto Sangyo University (KSU). We will then review the results of the 2009 entering cohort who were required to use ER with the 2008 cohort who did not. We will attempt to answer this question: Do students who do extensive reading outside the classroom as additional work achieve higher reading scores compared to students who are not required to do so.

Studies concerning the effectiveness of ER in an EFL environment have been limited. Both Krashen (2004) and Iwahori (2008) have surveyed prior studies, with virtually all of them indicating that cohorts experiencing ER show significant gains over classes using other methods. One important distinction that has not been included in these surveys, however, concerns the timing of the reading itself, which can be categorized as either *replacement* if the time spent reading replaces another classroom activity, or *additive* when the reading provides additional

contact time, normally outside of class hours. This paper is concerned with the effective implementation of additive ER.

Why Additive ER?

It should be utterly obvious that the more students can practice with the target language the better they will be able to use it. This not withstanding, relatively few teachers and even fewer school-wide curricula have implemented ER. As will be noted below, most empirical studies of ER in foreign language (FL) settings are based on replacement approaches. While many language educators acknowledge the value of extensive reading in a foreign language curriculum, factors such as book acquisition and management, student motivation and record keeping have conspired against its widespread adoption.

A single, enthusiastic instructor may go through the effort to bring books into class, stimulate the students to read outside of class time, and devise some form of record-keeping in order to hold students accountable for their reading. However, most teachers, even those who recognize the value of ER, do not have the free time or the stamina to implement an effective program. At the administrative level, it would be unfeasible to require such an implementation from already-overworked teachers.

Macalister (2010), for instance, in a survey of teacher attitudes to ER practices in New Zealand, suggested a host of measures that could encourage more ER *in the classroom*, but the abstract to the article fails to mention any additive approach:

> If extensive reading is to be more widespread in higher educational contexts, it needs to be promoted through teacher education, new research, changes in course design, raising awareness among administrators and managers, and improved resource provision (p. 59).

Teachers who are already saddled with more material to cover than feasible in limited classroom hours are reluctant to sacrifice time for ER in their classes (Takase, 2002). As mentioned in Helgesen (2005, p.26), “there simply is not enough class time for all that reading to happen during class time” (Helgesen, 1997, 2005).

Holding Students Accountable

If ER is conducted within the classroom, the problem of holding students accountable is virtually a non-issue since teachers can directly supervise the students. Similarly, if one is dealing with students who want to read outside of class time, no particular method of enforcement is required although feedback from students on their reading may serve other useful purposes in a language program.

The most common method of holding students accountable is for them to produce some form of report. Numerous variations exist, some of which are introduced in Bamford and Day (2004).

Others may be found in the various contributions on the http://extensivereading.net or http://er-central.com websites. Oral reports, reflections, group discussion and even picture drawing are reported in the literature (see Bamford and Day [2004] for numerous examples). All such approaches, however, require either additional work from the teacher or intrusion into already busy class activity schedules.

Our conclusion is that additive ER can only be widely implemented if there are effective means to hold students accountable for their work, that does not increase the workload of the teachers or unduly intrude on class time. MoodleReader discussed herein is one such method.

Literature Review

The Annotated Bibliography of Works on Extensive Reading in a Second Language (English Reading Foundation) lists over 500 articles related to the theory and implementation of ER. As is the case with much SLA research, it has been difficult to assess the effectiveness of ER. Studies conducted with ER within the classroom have been small in scale, lacked control groups or have been confounded by additional English classes concurrently with other instructors.

Studies of ER conducted outside the classroom are often confounded by their in-class study, making it difficult to attribute gains to the outside reading alone. Reseachers such as Stokes, Krashen, and Kartchner (1998), Renandya, Rajan, and Jacobs (1999), Mason (2006) and Lee (2007), have attempted to empirically demonstrate the effectiveness of reading large volumes of text as an aid to language acquisition. Each of these studies, however, was conducted on a small scale and suffered from some defects in the research design.

The studies listed below are relevant to the current study. See Iwahori (2008) for her review of these studies, which deals with some aspects not covered below. Table 1 lists a number of empirical studies utilizing a control group that have been concerned with the efficacy of ER when used in the classroom as a replacement for other classroom activities that were assumably deemed less worthwhile.

Table 1. *Studies on the efficacy of extensive reading in EFL*

Study	Total N Size	Milieu	Target Measure*	Reading Location
Robb & Susser (1989)	124	Univ., Japan		In/Out
Lai (1993)	266	Hong Kong Jr. H.S.	C, W & R	In
Masuhara et al. (1996)	91	Univ., Japan	C & V	In
Bell (2001)	26	Br. Council, Yemen	R	In
Sheu (2003)	98	Jr. H.S., Taiwan	C,V,G,R	In
Tanaka & Stapleton (2007)	190	H.S., Japan	C	In/Out

Note. *C: Comprehension; G: Grammar: R: Rate; V: Vocabulary; W: Writing

Robb & Susser (1989) compared a *Skills Approach* (SA) with an ER approach over the course of an entire academic year. The ER group read in class using material in *SRA Reading Laboratory Kits* (McGraw-Hill) and was required to read 500 pages outside of class from a library of readers

for American teenagers (no graded readers were available at that time). The SA approach used a standard reading skills book, with additional work assigned for homework. The ER group made significant gains in comprehension, vocabulary and reading speed compared to the SA group. The amount of outside work by the SA group, however, was approximately half the total time spent by the ER group. Thus the authors stated:

> Differences could have emerged purely because the EXTENSIVE group studied more. Had we been able to make the SKILLS group spend an equal amount of time studying (whether this is possible given the nature of the materials is another question), perhaps the results would have been radically different. It may be just that the EXTENSIVE readers are able to spend more time reading thanks to the nature of the material (p. 246).

Lai (1993) compared 266 students in an experimental 4-week summer program, with the results of an earlier year-long study. Her students read material of their own choosing for half of each class period and were allowed to take home books that they had not finished. Two of her three groups, which she termed as *motivated* made significant gains in comprehension and reading speed.

Masuhara, Kimura, Fukuda, and Takeuchi (1996) used two intact classes that were of differing ability from the onset, with the ER treatment group scoring lower on pretest assessments compared to the "Skills Training" group. Despite the initial differences in ability, the ER "Reading Experience" group had caught up to the skills group by the end of the 8-week study.

Bell (2001) reported on a study conducted with young adult students working in various government ministries in Yemen Arab Republic:

> It measured both reading speeds and comprehension in two groups of learners exposed to 'intensive' and 'extensive' reading programs respectively. The 'extensive' group was exposed to a regime of graded readers while the 'intensive' group studied short texts followed by comprehension questions. Results indicate that subjects exposed to 'extensive' reading achieved both significantly faster reading speeds and significantly higher scores on measures of reading comprehension" (from the abstract of the article).

Sheu (2003) used two experimental groups, one reading graded readers and the other books written for native speaking children at a junior high school in Taiwan, for a period of one academic year. A third group formed the control. Reading was done only in class for one 45-minute session per week. Both experimental groups out-performed the control group significantly in reading comprehension and grammar assessments, but only the graded reader group showed progress in vocabulary, presumably due to the more complex vocabulary in the native speaker texts.

Tanaka and Stapleton (2007), in an experiment in a Japanese high school, demonstrated that their "treatment group, especially those who read graded readers, scored significantly higher in reading speed and comprehension than the control group" that did no extensive reading (p.115). They conclude that "Japanese high schools and more broadly, English teachers in input-poor EFL settings should increase reading input within the students' linguistic levels both inside and

outside of the classroom" (p. 115).

The current study

Two other issues cloud the past research. Most studies have been conducted by a single instructor, usually the author of the paper, which brings into question the role of bias. Ethical reasons have also limited the scope of experimental studies, since those who believe in the value of extensive reading would not want to withhold it from a subset of their students in order to obtain a control group.

This study overcomes these two obstacles. It avoids the question of teacher bias since all classes in the 2009 academic year were required to undertake an ER program. It avoids the ethical question since two full entering cohorts are compared, the first cohort having had no ER since it had not yet been introduced.

The ER program was facilitated by a software program integrated with the Moodle course management system, which allowed implementation across a curriculum with some 70 instructors, with little need for teacher training.

Brief description of the MoodleReader program and its implementation

MoodleReader is an add-on, developed by the first author, to the standard Moodle course management system. The module allows students to take short, timed quizzes with randomized questions on their graded reading, controlling the level of book for which they can take quizzes and the frequency with which they can access them. The module was developed as a reaction to the inadequacies of the Accelerated Reader Program (Renaissance Learning), which had been in use for several years with the students majoring in English at Kyoto Sangyo University. A thorough description of the module is beyond the scope of this paper. See Robb (2010) or visit http://moodlereader.org for a detailed description.

The module was launched in time for the 2008–2009 school year solely for the first-year student English majors, but was then expanded to include virtually all non-major 1st year students in the following year. This report is based on a comparison of the non-major cohort in the 2009–2010 school year, comparing them to the same population of students the previous year who followed the same curriculum but without the benefit of an extensive reading component and thus served as a control group. Although the two cohorts were not studying simultaneously, they were initially judged as equivalent as measured by identical items on the KSU placement test that is administered to the entire student body on April 2 each year, the second day of the academic year. This baseline measurement is explained in the following section.

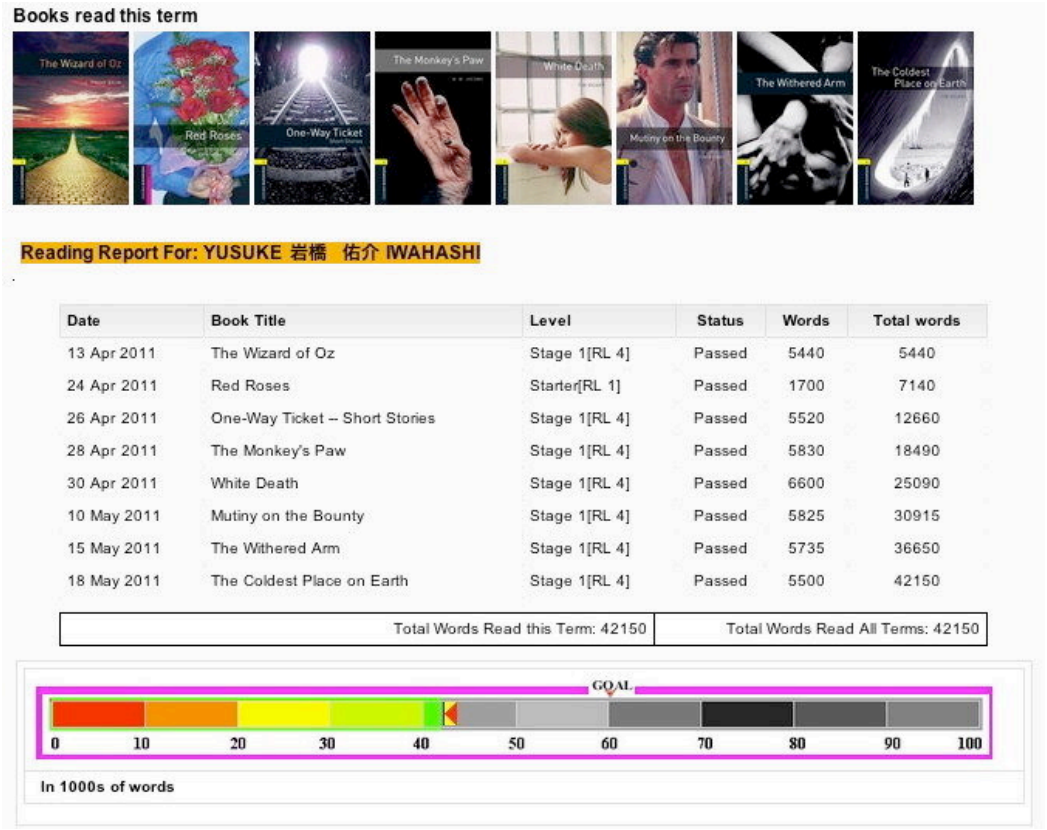

Figure 1. The MoodleReader Student Screen

The English curriculum at Kyoto Sangyo University

All non-major first year students at Kyoto Sangyo University are required to take eight credits of language study during their first year. The choices are listed below, along with the number of students in the 2008 and 2009 academic years who elected each combination. Each two-credit course meets twice a week, for 15 weeks per term. Each class meets for 90 minutes per meeting.

Table 2. *Distribution of language study choices (2008 and 2009)*

	2008	2009
Oral Communication and Reading Skills	872	430
Oral Communication and another foreign language	1131	1097
Reading Skills and another foreign language	596	693
Eight credits of another foreign language, no English	198	366
Total:	2797	2586

As shown in Table 3, students are divided into five levels based on the placement test with approximately 40% of the students being placed in the middle level. Level 1, the lowest, and Level 5, the highest have approximately 10% each of the total enrollment.

Table 3. *Students enrolled in each level for the 2009 academic year*

Level 1	Level 2	Level 3	Level 4	Level 5	Total
204	438	971	406	201	2220

Note that the total number is considerably less than the number of entering students who took the initial placement test. This is because the Faculty of Culture, as well as the English and International Relations departments of the Faculty of Foreign Languages, have a separate curriculum. Furthermore, as shown in Table 2, approximately 200-300 students selected the eight-credit foreign language option and are thus not part of this study. For reference purposes, we estimate that the Test of English as a Foreign Language (TOEFL) scores for these levels would range from the very bottom of the scale to perhaps 430-450 on the average for Level 5, so the entire student body can be classified as *basic* or *lower-intermediate* in English language ability.

The curriculum offered to the students in 2008 and 2009 was essentially identical. The students were divided into a total of 126 sections in 2008 and 127 in 2009. The instructors selected their text from a short list of approved textbooks for each level. There was very little turnover in the instructors between the two years. Of the 59 instructors participating in the Oral Communication (OC) and Reading Skills (RS) courses, 44 had taught in both 2008 and 2009. Thus the only major factor that distinguished 2009 from 2008 was the required extensive reading in 2009.

Hypothesis

The students in the 2009 cohort will have significantly higher reading scores than the 2008 cohort despite being of equivalent ability at the outset, as measured by the KSU final examination.

Baseline comparison

Kyoto Sangyo's General Education English Center has developed a series of three 60-minute examinations, the first of which is administered as a placement test, while the other two are used as final examinations for the Spring and Fall terms, with between 2500 and 3000 examinees for each administration. The tests contain a reading section with 32 questions, administered over 35 minutes, and a listening section with 36 items that takes approximately 20 minutes. All three examinations were originally created from a single pool of items, to which a Rasch analysis had been applied. The best performing items where then evenly distributed according to their difficulty among the three sets. The exams have undergone further tweaking after each annual administration so they cannot be claimed to be perfectly equivalent. In our analysis we used a simple measure of correctly answered items.

The placement test had undergone revision between April 2008 and April 2009, with some reading passages completely replaced and other items tweaked, which left only 10 unchanged items that could be used to establish baseline equivalence between the two years. These items showed good discrimination, and considering that the N-size was over 2500, they were

considered sufficient to establish the equivalency of the two cohorts.

Table 4. *Scores on 10 identical reading items (April 2008 & April 2009)*

	2008		2009			
Faculty	Average	*N*	Average	*N*	t-tests	*sig.*
Business	5.3940	713	5.2385	647	0.0874	ns
Economics	5.3472	651	5.0423	632	0.0076	p<0.001
Law	5.1820	629	5.2110	704	0.8022	ns
Foreign Languages	5.9667	487	6.0232	479	0.6853	ns
Science	4.7168	115	4.7547	106	0.8910	ns
Technology	5.4848	66	5.3644	52	0.6556	ns
Computer	5.0753	151	5.1801	163	0.6595	ns
ALL	5.4063	2812	5.3990	2783	0.7714	ns

Comparison was done faculty by faculty since there can be significant fluctuations from year to year in specific faculties due to such factors as alterations in the curriculum, the addition of a new department within a faculty, and conflicts with the entrance examinations of other universities which either depletes or enriches the applicant pool. While multiple t-tests are frowned upon in some statistical circles, the utter insignificance of the t-test values, except for Business which approached the 0.05 level and Economics which was highly significant condones its application here. Since the object of this study is to demonstrate that the application of extensive reading in the 2009-2010 year had a significant effect on reading ability, the fact that the averages for both the Business and Economics faculties were lower in the experimental condition should contribute towards lowering the average scores on the final examinations and thus, make it more difficult to demonstrate a significant difference between the two years. As we shall see, this notwithstanding, there was significant improvement.

Procedure

All students taking either the OC course or the RC course were required, as part of the course syllabus, to read five graded readers outside of class time. Students registered in both courses were required to read a double dose. In later years, the reading requirement was set in "total words read" rather than "number of books read" in order to even out the approximate time investment required for students at varied levels of book difficulty and length.

For their grade in the course, we simply said that their grade—based on their teacher's evaluation of their class performance (70%), and the final examination (30%)—would not change if they achieved their reading target, but that five points would be added to their grade if they doubled the requirement, and, conversely, they would lose five points if they did nothing, with their scores scaled, of course, between these two extremes.

Since more than 50% of the teaching staff was primarily composed of part-time teachers who

were rarely on campus except for the time when they were actually teaching, it was not feasible to train the teachers in the ER approach. Instead, we merely requested the teachers to distribute a handout in their first class that explained in Japanese the purpose of ER, their reading requirement, how to borrow books, how to access and take quizzes using the MoodleReader system and how they were to be evaluated. The results at the end of each term were distributed to each instructor along with each student's grade on the final examination for incorporation into their final evaluations.

Results after two terms of study

This section analyzes the results of the alternate version of the same test set that was used in both years as the final examination in February 2009 and February 2010. This time, an identical set of questions was used for the entire examination, with 32 Reading items and 34 Listening items.

Table 5. *Comparison of results by faculty after two terms of study*

	February 2009 (No ER)			February 2010 (ER)			ANOVA	
	Reading	Listening	N	Reading	Listening	N	Significance	
	Avg. (SD)	*Avg. (SD)*		*Avg. (SD)*	*Avg. (SD)*		Reading	Listening
Economics	11.59 (3.9)	16.72 (5.2)	555	15.64 (4.2)	17.39 (4.0)	543	0.000 *	0.016
Business	12.06 (4.2)	15.96 (4.7)	675	16.02 (4.4)	16.92 (3.9)	587	0.000 *	0.001 *
Law	11.00 (4.3)	16.00 (5.7)	512	15.56 (4.2)	16.87 (4.0)	546	0.000 *	0.004 *
Foreign Languages	11.33 (5.1)	16.54 (6.9)	281	16.00 (4.6)	18.77 (5.1)	254	0.000 *	0.000 *
Science	10.75 (4.2)	16.04 (6.0)	111	15.07 (4.8)	17.07 (4.6)	100	0.000 *	0.131
Technology	11.90 (4.1)	17.65 (4.8)	63	14.81 (4.8)	17.00 (4.0)	48	0.001 *	0.490
Computer	10.64 (4.0)	16.55 (5.4)	145	15.86 (3.8)	16.63 (3.7)	142	0.000 *	0.888

Note. * $p < 0.01$

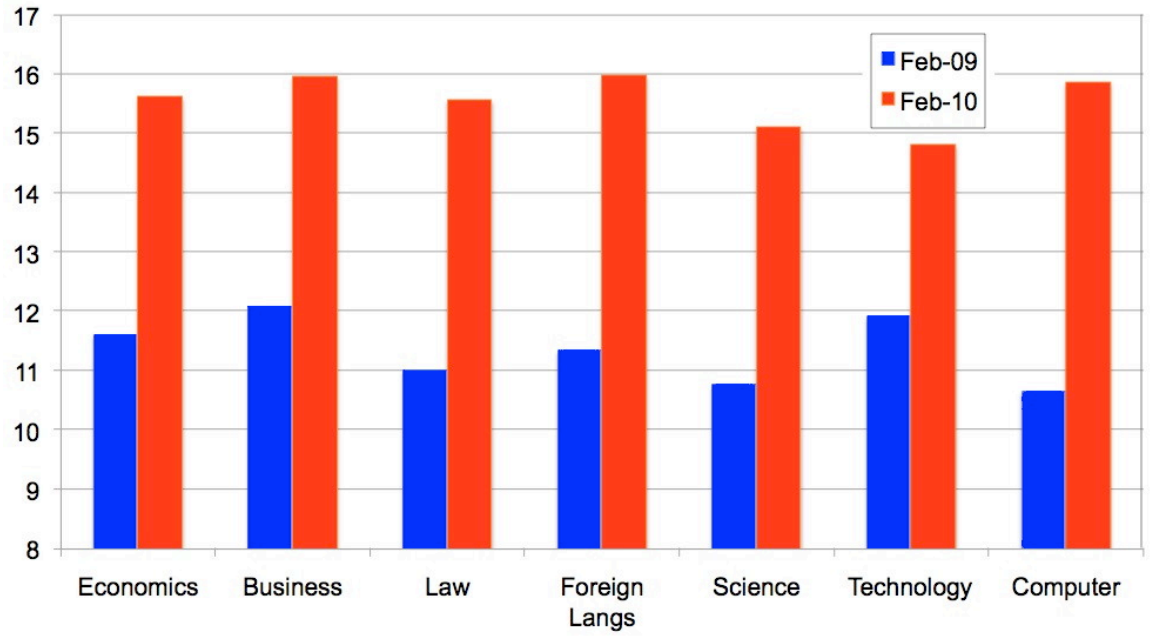

Figure 2. *Results by faculty (Reading)*

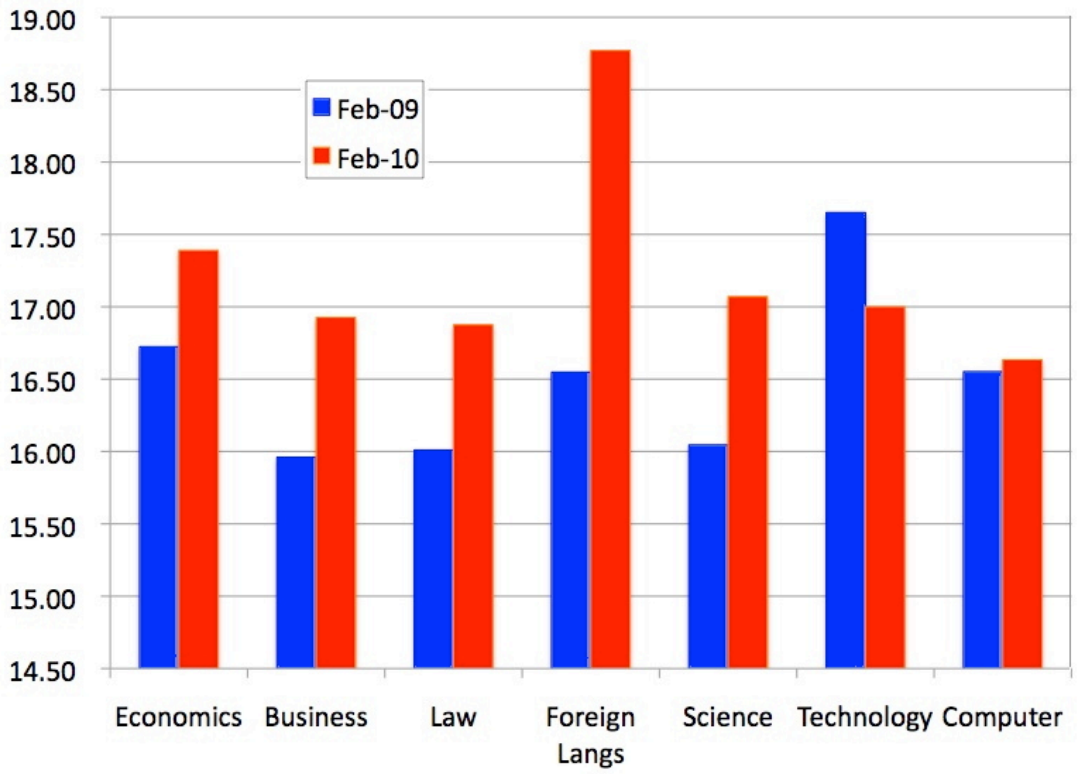

Figure 3. *Results by faculty (Listening)*

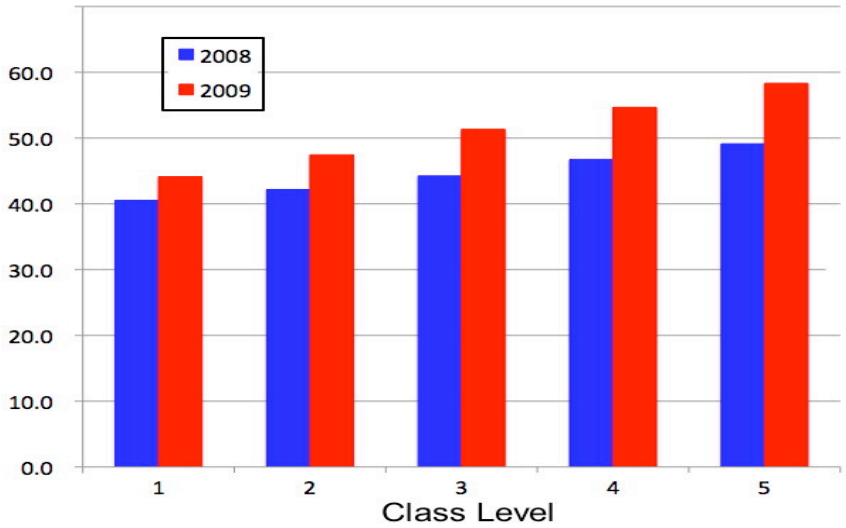

Figure 4. *Reading scores for the two years compared by class level*

As can be observed from Table 5 and Figure 2, the average score for students of every faculty showed a significant gain on the reading section compared to the average scores of the 2008 cohort. Three faculties also demonstrated significant gains on the listening section, as well, compared to the 2008 cohort (Table 5 & Figure 3). Figure 5 displays the reading data, but this time grouped by the class levels, explained earlier. This reveals that the Level 1 students in 2009 performed approximately as well as the 2008 students in Level 3, while the Level 3 students performed better than those in Level 5 of the previous year.

Discussion

It is clear from the data that the 2009 cohort which was required to read extensively did indeed improve compared to the 2008 cohort, and that this improvement can be solely attributed to their extensive reading. This does not, however, demonstrate that it was the extensive reading approach *per se* that caused the improvement. An alternate explanation might be that the improvement can be attributed to the additional contact hours that the students spent with English. It is conceivable that any other activity which required the students to actively use the language for an equivalent number of hours might have resulted in similar gains.

While this may be true, there are no other methods known to us which might be able to force a large number of students who are generally unmotivated to study English to spend the requisite amount of "quality time" with the language outside of class.

While motivated students will gladly do whatever they can to improve their language proficiency, other students will normally not do outside work unless there is a means for their instructor to verify that their work has been done.

Macalister (2010) acknowledges this quandary, but comes out solidly for a replacement rather than an additive approach: "students are 'not keen on doing work in their own time, many don't read in their first language.' If this is seen as an issue, then it is reasonable to suggest that

teachers should be considering solutions, with one solution being, clearly, the devotion of at least some in-class time to actually reading for pleasure" (p. 69).

ER, when conducted by a dedicated instructor with a small number of students, who can take a personal interest in each student's reading, may achieve greater overall success. It is likely that any method aiming to be applied across an entire curriculum will not enjoy the enthusiasm and support of the entire cohort of busy instructors. We have found, however, that many of our instructors, after observing the effect of ER on their students, have become converts to the approach, actively encouraging their students to read more with a concomitant greater improvement in their overall proficiency.

We believe that the implementation of ER at Kyoto Sangyo University has been successful because of a number of factors: (a) The administration was able to require the implementation of ER over the entire General English curriculum; (b) MoodleReader program relieved the instructors of the need to confirm that their students had done the requisite reading; (c) the final grade for their reading was handled by the administration, which simply reported the grades to each instructor via an Excel document along with the grades on the final examination. Teachers could easily insert their own class grades into this Excel document and calculate final grades using the built in formulas; and (d) the university library was supportive and willingly purchased the requested number and variety of graded readers, displaying them to the students with color-coded labels representing their general level of difficulty.

Conclusion

The results achieved in 2009 were attained with only about 70% of the students having taken quizzes on one or more books. (Five books was the requirement.) We can hypothesize that if more students had indeed "cooperated," the rise in scores would have been even greater. Naturally, there is a limit to the proportion of students who can be expected to conform to any requirement for additional outside study, but that proportion will increase in direct proportion to the weighting of their outside reading in their final grade. In a future paper, the authors will report on how a change from the +/- 5 points system outlined above, to one where a full 20% of the grade depended on their ER work, positively influenced the proportion of students conforming to the requirement.

In summary, we can say that an ER program in an additive mode, can be successfully implemented curriculum-wide if all of the following conditions can be met:

1. The administration requires ER from all students in a specific range of classes.
2. There is an effective way to hold students accountable for their reading that does not increase the instructors' workloads.
3. Likewise, final assessment is performed in a manner that is relatively trouble-free for the instructors.
4. Book management is handled in a centralized, efficient manner, through the school library or a self-access center.

References

Bamford, J. & Day, R. R. (2004). *Extensive reading activities for teaching language*. Cambridge: Cambridge University Press.

Bell, T. (2001). Extensive reading: Speed and comprehension. *The Reading Matrix*, *1*(1).

Extensive Reading Foundation. The annotated bibliography of works on extensive reading in a second language. Retrieved from http//:erfoundation.org/bib/biblio2.php

Helgesen, M. (1997). What one extensive reading program looks like. *The Language Teacher*, *21*(5), 31–33.

Helgesen, M. (2005). Extensive reading reports: Different intelligences, different levels of processing. *Asian EFL Journal*, *7*(3), 25–33.

Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language*, *20*, 70–91.

Krashen, S. D. (2004). *The power of reading: Insights from the research (2nd ed.)*, Portsmouth, NH: Heinemann.

Lai, E. F. K. (1993). The effect of a summer reading course on reading and writing skills. *System*, *21*, 87–100.

Lee, S. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal*, *38*, 150–170.

Macalister, J. (2010). Investigating teacher attitudes to extensive reading practices in higher education: Why isn't everyone doing it? *RELC Journal*, *41*, 159–175.

Mason, B. (2006). Free voluntary reading and autonomy in second language acquisition: Improving TOEFL scores from reading alone. *The International Journal of Foreign Language Teaching*, *2*, 2–5.

Masuhara, H., Kimura, T., Fukada, A., & Takeuchi, M. (1996). Strategy training or/and extensive reading? In T. Hickey & J. Williams (Eds.), *Language, education, and society in a changing world* (pp. 263–274). Clevedon, UK: Multilingual Matters.

Renaissance Learning, Accelerated Reader (software), http://renlearn.com

Renandya, W. A., Rajan, B. R. S., & Jacobs, G. M. (1999). Extensive reading with adult learners of English as a second language. *RELC Journal*, *30*, 39–61.

Robb, T. (2010). A digital solution for extensive reading. In R. R. Day (Ed.) *Bringing extensive reading into the classroom* (pp. 105–110). Oxford: Oxford University Press.

Robb, T. & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, *5*, 239–251.

Sheu, S. P-H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter*, *36*(2), 8–26.

Stokes, J., Krashen, S., & Kartchner, J. (1998). Factors in the acquisition of the present subjunctive in Spanish: The role of reading and study. *ITL Review of Applied Linguistics*, *121–122*, 19–25.

Takase, A. (2002). The effects of extensive reading on Japanese high school students. *JALT 2002 Conference presentation* summary.

Tanaka, H. & Stapleton, P. (2007). Increasing reading input in Japanese high school EFL classrooms: An empirical study exploring the efficacy of extensive reading. *The Reading Matrix*, *7*(1), 115–131.

About the Authors

Thomas Robb is a professor at Kyoto Sangyo University where he is Chair of the Department of English, Faculty of Foreign Languages. He is a long time user of CALL and the Internet, and has created a number of websites for various student projects, interactive learning and professional exchange. He is currently Lead Developer for the MoodleReader software. E-mail: trobb@cc.kyoto-su.ac.jp

Makimi Kano is an associate professor in the Faculty of Cultural Studies at Kyoto Sangyo University, Japan. Her research interests include corpus-based lexical research, lexicography, extensive reading, extensive listening, and vocabulary acquisition. E-mail: makimik@cc.kyoto-su.ac.jp