The Reading Matrix © **2010**
Volume 10, Number 2, September 2010

Examining Multiple Variables within a Single ER Setting

Omar Karlin
Meiji University

Rick Romanko
Wayo Women's University

ABSTRACT

This study examined the gains in student affect, vocabulary, and reading fluency for 110 university students in an extensive reading program in Japan. It was important to measure all of these dimensions within a single study and teaching methodology, so gains could be appropriately compared against each other. The adopted teaching methodology was a communicative one which stressed a number of in-class activities with out-of-class reading, with reading speed, vocabulary, and comprehension measured over the course of a semester, and paired sample t-tests were conducted using pre- and post-test scores on six variables. Students were also clustered in higher-level and lower-level groups to determine if they differed in their learning rates. Results indicated that affect increased substantially, while fluency increased minimally, and vocabulary did not increase at all. Paired sample t-tests indicated that the lower-level students gained more in terms of fluency than the higher level students.

INTRODUCTION

For English learners, especially those in an EFL context where access to an L1 community is constrained and viable language input is minimal, there are a number of benefits realized with *extensive reading* (ER). Carrell and Carson (1997) define extensive reading as involving "rapid reading of large quantities of material or longer readings (e.g., whole books) for general understanding, with the focus generally on the meaning of what is being read than on the language" (p. 50). One of the central tenets of extensive reading is the potential boon to affective dimensions, such as motivation, confidence, and enjoyment. With respect to research done in Japan, a number of studies have shown ER-related benefits to effect, specifically, university-student motivation (Robb & Susser, 1989), high-school student motivation (Hashimoto, Takada, Isobe, Sakai, Ikemura, & Yokogawa, 1998), and even teacher motivation (Takase, 2006). While not enjoying as much of a consensus in support, vocabulary-learning is another oft-cited benefit. There has been some disagreement over the amount of vocabulary that is actually learned through ER, from more optimistic proponents (Mason & Krashen, 1997) to more tempered views (Waring & Takaki, 2003). However, it is worth noting that even though some (e.g., Nation, 2001; Waring & Takaki, 2003) have found fault with the lack of rigor in the more optimistic studies, critics of overzealous research have contended that ER can do a lot to help strengthen existing vocabulary knowledge, as pointed out by Waring and Takaki (2003): "it is our contention that ultimately learners do not learn a lot of *new* words from graded reading, but in fact graded reading helps to deepen and consolidate *already known* language" (p. 154). One final

benefit to be mentioned in this brief outline is the effect ER has upon reading fluency. The nature of ER, in which repeated exposure to letters, words, and even texts is maximized, is ideally suited for creating reading automaticity (Logan, 1997) and this has been borne out in the research (Taguchi, Takayasu-Maass, & Gorsuch, 2004). In a study that focused on Japanese high-school students, ER was found to significantly improve reading fluency (Iwahori, 2008).

Suffice it to say, research has been supportive, in varying degrees, of affective, lexical, and reading fluency benefits catalyzed by ER. However, one area that has been lacking is the examination of these multiple variables within a single study and teaching methodology. It is difficult to take research at face value when ER practitioners often rely on vastly different teaching approaches. Some adhere to a hands-off approach in which students (and sometimes the teacher) are encouraged to read within the class and are not saddled with reports, assignments, or other burdensome activities that might sap the intrinsic development of a reading habit (Day & Bamford, 2002). Others feel that a more comprehensive approach that integrates skills and recycles concepts through a variety of assignments is of immense benefit to students (Hunt & Beglar, 2005). If one study claims a vocabulary result with the former methodology, and another claims a motivational benefit with the latter methodology, questions should be raised as to whether the results are commensurate**.** In fact, conditions that spur vocabulary development *might* rely on rigid recycling of vocabulary through book reports and vocabulary-isolating activities, while a study that is intent on fostering affective gains, may intentionally minimize such conditions, thereby making these results problematic when cited as uniform benefits of ER. It is therefore crucial to view all of these potential benefits in a unified research design, not only to eliminate potential contradictions, but also to give insight as to the degree that certain phenomenon, such as increases in vocabulary acquisition and fluency can occur in tandem.

Another area of ER that may be under-researched, at least in relation to the aforementioned benefits of affect, vocabulary, and fluency, is how these benefits are realized by the different students within a class. Clearly, not every learner is the same, and the benefits of ER weigh upon different learners in different ways. In an interesting study on de-motivation factors that impede poorly motivated ER students, Takase (2003) outlined the obstacles separating the highly-motivated from the poorly-motivated. In another study, Mori (2004) identified several predictors that manifested in different reading intensities in ER students, implicitly distinguishing between different student types. While these studies examined the motivation and reading intensity of different types of students in an ER class, they did not focus on linguistic variables such as vocabulary knowledge and reading fluency. This study proposes to identify class-wide benefits of ER, identify the different types of students in an ER class, and contrast the benefits realized by these different types of students.

CLASS-WIDE AND GROUP BENEFITS

As already mentioned, there are various benefits to ER. With this study lasting only a semester, it was thought that some of these benefits could be realized, while others might not. For example, the novelty of ER was expected to have an immediate impact upon students' affective perceptions of their own ability. Very different than what most Japanese students have experienced in previous schooling (Gorusch, 2001), as well as being designed to enhance students' confidence because of its ease (Ono, Day, & Harsch, 2004; Takase, 2004), ER affective benefits are likely to be quick and substantial. Perhaps not as dramatic, but still significant, would be fluency gains. The initial transition for students into a reading habit will most likely have a considerable impact upon their processing speed. However, it is likely that this improvement would stabilize after the initial increase, possibly yielding diminished gains over time. Finally, vocabulary knowledge was not expected to improve within the short timeframe of this study.

Opinions about vocabulary gains arising from ER have been mixed at best, with a number of key studies showing little or no gain. Presumably a single semester would not be enough to realize any significant gains. Ultimately, if this research does indeed show benefits to only student affect and fluency, it may underscore the need for curriculum designers to carefully think about how ER is being used within their curricula. If their goal is to have students feel better about studying English, then isolated course offerings of ER should be sufficient. However, if more ambitious goals are sought, such as improving TOEFL scores or boosting linguistic capabilities, then a curriculum-wide commitment to ER, extended over more than just a single semester, may be necessary in order to yield more than affective results. Of note here is that the teaching methodology used in this study had a heavy focus on skill integration within the class through communicative activities and independent reading outside of class. Some ER practitioners may dispute this approach because it deemphasizes the reading focus in favor of a more balanced approach. For the purposes of this study, a single teaching methodology was used to establish a baseline of findings for future research in which competing teaching methodologies may be compared. At present, understanding how affect, vocabulary, and fluency function within a single context, and their functions in relation to each other, is highly advantageous.

Further, it would also be advantageous to know how different kinds of students benefit from ER. It was thought that a cluster analysis of the participating students would result in a variety of clusters, presumably based on proficiency (for example, beginner, intermediate, and advanced levels). In this particular study, proficiency would refer to a composite of the variables being measured (i.e., affect, vocabulary, and fluency). In all, six variables were included in the cluster analysis: perceived vocabulary, actual vocabulary (both productive and receptive), perceived reading ability, perceived overall English ability, and reading speed. If the prediction of proficiency-separated clusters is validated via their post-test scores, such clusters might indicate different patterns of program success. Presumably lower-proficiency students, who have more negative associations with studying English, will see the greatest benefits in affect. These learners might also see strong benefits in fluency since they are beginning from a low starting point and have a higher potential for improvement. If any of the clusters are to show vocabulary gains, this lower-level cluster, which is based on an amalgam of proficiency variables, seems the most likely since a significantly more positive learning experience may spur increased effort at learning new vocabulary. Conversely, higher-level students may see more muted gains because of a higher starting point in all variables. Should these predictions be confirmed, it may indicate that ER is more effective with lower levels, while higher levels might benefit more from an intensive reading approach (although additional research with a more contrastive framework between intensive and extensive reading would need to be conducted to confirm this).

Research Questions

Examining the three dimensions of affect, vocabulary, and reading fluency in conjunction provides a comprehensive view of the effects of ER and its overall benefits for students. Similarly, examining various groups of students within an ER program may provide insights as to the best utility of ER courses within an established curriculum. This study will attempt to answer the following three hypotheses:

1. Participants, as a whole, will see a significant improvement in affect and reading fluency, but not in vocabulary knowledge.
2. Student affect will show more substantial gains than fluency, and both will show more significant gains than vocabulary (if there are any).
3. Clusters of students will experience different degrees of success in the ER program, as evidenced by post-test scores.

Research Design

Participants

Participants in this study initially included 116 first- and second-year students at a national university in Tokyo, Japan. All participants were enrolled in a compulsory English reading course in which the medium of instruction was English. Within the sample, 47 of the students belonged to the Faculty of Agriculture while the remaining 69 students were from the Faculty of Technology. Students represented a number of different majors from the two aforementioned faculties. Some participants were receiving other English instruction within the university, although most were not, and those that were receiving other instruction were enrolled in no more than a single additional course. Three students had incomplete data, and three other students were statistical outliers, resulting in all six of them being dropped from the study. The decision to eliminate the statistical outliers from the study was based on the disproportionate influence their statistical data would have had on the entire data set. In the case of outliers exerting an undue influence on a data set, it has been suggested that they be eliminated entirely (Tabachnick & Fidell, 2007). The final number of participants was 110 (85 male, 25 female). Participants came from three separate classes which were assembled based on their major and/or faculty, and all participants exhibited approximately the same low-intermediate to intermediate language proficiency. While level testing was not included as part of this study, to assist the reader in understanding the approximate level of the students' language proficiency, the researchers would estimate that TOEFL scores (Internet-Based Test) were in the 40-60 range, while TOEIC scores would be in the 400-600 range, although this is only an estimation and not supported by documented test scores.

Procedure

The first meeting of the course was used as an orientation class in which students learned about the course syllabus, rules, and philosophy of the class: *people learn to read by reading and the best way to become a better reader is through reading, reading, and more reading*. It was explained to them that they would be expected to read at least ten books of their choice (about a book a week) for pleasure during the semester. The length of each book ranged from 15 to 85 pages. On average, students read books that were about 35-45 pages long. There were approximately 600 graded readers available to students drawn from two publishers, Penguin and Oxford, and ranged in level from Easy Starts, containing up to 200 headwords, to Level 6, containing up to 3000 headwords. By the end of the semester, students had each read from 178 to 748 pages, with a mean average of 397.49 pages. All reading was done on students' own time outside of class. Each week, students were asked to complete a reading tracking sheet to indicate how many pages they had read that week. To ensure that students would not record a fraudulent number of pages read, they were reassured that a low number of pages read would not negatively effect their course grade. Additionally, random students were selected by the teacher each week for a short oral interview, in which the teacher asked pointed questions about the student's graded reader to confirm that the students actually did the stated reading. During the second meeting of the course, a series of questionnaires and tests were administered to students in order to measure their affect, vocabulary, and reading fluency. Similar tests were administered again during the final meeting of the semester.

Affect

With regards to affect, a 13-item questionnaire was created based upon self-efficacy principles which Dornyei (2005) defined as "one's beliefs in one's capabilities to carry out certain specific tasks" (p. 213). The questionnaire was arranged in two sections. The first section (ά = .815), comprised of six items, focused on student self-perceived reading ability, while the second section (ά = .894), comprised of seven items, focused on self-perceived overall English ability. As for face validity, the items were shown to two other native English professors to confirm that the items were appropriate for this type of research.

Students answered items on a four-point Likert scale that ranged from strongly negative to strongly affirmative. Students were asked to identify themselves on their questionnaires, but were assured that the questionnaires would not influence their grade. To ensure complete understanding of the questionnaires' instructions by all students, regardless of their English proficiency levels, the instructions were written in both English and Japanese, following Brown's (2001) questionnaire design principles. The instructions were translated into Japanese by a bilingual non-native English professor at the university. Simple and easy-to-understand English was used for all questionnaire items. It was believed that if the questionnaire items were written at an appropriate level of English, such items would not hinder students' comprehension. Two other native English professors also looked at the items to form a consensus that the level of English was appropriate (Brown, 2001), and well within students' range. The exact wording of the questionnaire items can be seen in Appendix A. As an additional safety measure, during the administration of the questionnaires, students were allowed to use their dictionaries for any words that they did not understand. In addition, the teacher was also available to answer any questions the students may have. For the statistical analysis, each section of the questionnaire was totaled to form a composite score, one relating to reading ability and the other relating to overall English ability.

Vocabulary

For vocabulary assessment, students completed Nation's 2000-word level and 3000-word level productive (Nation & Laufer, 1999) and receptive (Nation, 1990) vocabulary tests. There was a noticeable *floor effect* involving the 3000-word level productive and receptive tests (i.e., all of the students were scoring poorly on it), so the 3000-word tests were dropped from the statistical analysis**.** In addition to the 2000-word level tests, students were given a 2000-word level perception test in which they were asked to estimate for themselves how many words they did not know. This was done by simply giving each student a copy of West's (1953) General Service List of English Words, which is the classic list of 2000 high-frequency words of English, and asking them to circle the words on the list that they did not know. (A slightly adapted version of the original list can be accessed at http://www.nottingham.ac.uk/~alzsh3/acvocab/wordlists.htm). The intent of this test was to provide some insight as to the students' *perceptions* of their vocabulary (not their *actual* vocabulary). Since it was a test of words they did not know, a lower score actually meant an improving vocabulary (this is important to remember when viewing the statistical results). In all, three tests were included in the statistical analysis, Nation's 2000-word level productive and receptive tests (assessing actual vocabulary knowledge) and a 2000-word level perception test (assessing perceived vocabulary knowledge). However, the 2000-word level perception test was not considered an indicator of vocabulary knowledge because it lacked rigor: Students were only asked if they felt they had learned a word, with no way to ensure that their impressions were accurate. Students were not required to demonstrate their understanding by placing vocabulary in appropriate sentences. The result was an improvement in acquired vocabulary that would have greatly exceeded even the most optimistic proponents of vocabulary

acquisition through ER. It was thought that this occurred because students may have forgotten words that they actually knew when doing the pre-test and, furthermore, may have been overly confident by the end of the semester by claiming to know words they did not. Nonetheless, this was considered an affective test precisely because it measured students' perceptions of their own development and competence.

Reading Fluency

With regard to reading fluency, students were asked to read a text passage at a comfortable speed. Two actual tests were used at each testing phase, with half the subjects completing Test A and the other half completing Test B. During the final testing period, the tests were swapped, with each half of the subjects completing the test that they had not yet seen. Before the tests were administered, the texts used in each of the tests were analyzed using a Web Vocabprofiler (Cobb, 1994; Heatley & Nation, 1994) to ensure students would be able to read the texts fluently. Results showed that the tests were 273 and 289 words long respectively, and that the lexical frequency coverage was 91% (K1) and 96% (K1 + K2) for Test A, and 94% (K1) and 97% (K1 + K2) for Test B. These coverage figures reflect the removal of country nouns, regional nouns, and names from the analysis. The results satisfy Hu and Nation's (2000) suggested 95% coverage of the running words in a text to ensure fluent reading. The texts were adaptations taken from the Oxford Dominoes reading series (Escott, 2002; Fitzgerald, 2002) and can be seen in Appendix B. When reading the assigned passage, students were instructed to not take too much time, but also warned against racing through it since there would be a small comprehension test after they finished (to ensure that they did not recklessly speed-read). The number of words in the passage was divided by the amount of time taken to read the passage, to establish a *words-per-minute* (WPM) score. A comprehension test of five questions was administered after students finished reading to ensure they abided by the rules of the activity. All of the students passed the comprehension tests, scoring either 4 or 5 out of five. To represent reading fluency, a WPM score was used in the statistical analysis. This measure of fluency was modeled after an already established test (Nuttall, 1996). Additionally, a single mid-semester fluency test was also administered to all students. The text was 303 words long, and the lexical coverage was 91% (K1) and 93% (K1 + K2). Of the three tests administered, this mid-semester test had the least lexical coverage.

Statistical Analyses

As stated earlier, of the 116 participants, three were removed from the study because they had incomplete data (specifically, missing pre-test vocabulary data), and another three were removed because they were statistical outliers. With the remaining 110 students, six paired-sample t-tests were conducted. Each paired-sample t-test was based on a pre- and post-test variable. The six variables were: (1) a 2000-word level perceived vocabulary test, (2) a 2000-word level production vocabulary test, (3) a 2000-word level reception vocabulary test, (4) a perceived reading ability questionnaire, (5) a perceived overall English ability questionnaire, and (6) a WPM score. For both the initial class-wide analysis and the subsequent cluster analysis, a one-tailed hypothesis was selected since the literature suggests that affect and fluency would significantly improve, while vocabulary would not. Also, post-hoc correction methods were conducted using Holm's (1979) sequential procedure. It was thought that Holm's procedure would have the statistical power to avoid Type 1 errors, yet be flexible enough to also prevent Type II errors, as is evident in the second analysis involving different clusters of students.

Following the class-wide paired-sample t-tests, a hierarchical cluster analysis was conducted in order to determine the most appropriate number of clusters in which to divide the

students. The selected cluster variables were the 2000-word level perceived vocabulary pre-test, the 2000-word level production and reception pre-tests, the reading ability and overall English ability questionnaire pre-tests, and the WPM pre-test. It was thought that these six variables might group students in unexpected ways, such as slow and deliberate readers with a comprehensive vocabulary, or speed readers with a smaller vocabulary, or strong readers with poor motivation, or weak readers with high motivation. Ultimately, clusters were based on a simple high versus low orientation which seemed to justify using a reading proficiency test to establish groups, instead of going through the trouble of a cluster analysis. However, it was not clear at the outset that these six variables would simply cluster into simple high and low orientations. If one believes that readers can be characterized as more than just "high" or "low," then a cluster analysis would be the most appropriate course of action.

Only pre-tests were included in the cluster analysis for fear that if post-tests were also included, it would have been impossible to monitor the progress of clusters over the course of the semester. If post-tests were included, clusters would have been determined by statistical factors that occurred during the semester, while the goal of this study was to monitor progress over the semester (which required groups to be determined before the ER treatment was administered). Ward's method (Kettenring, 2006) was selected as the cluster linkage method because other linkage methods produced chaining in the resulting dendrograms, as seen in Figures 1 and 2. Ward's linkage, however, yielded a dendrogram with clearly discernable clusters, as seen in Figure 3.

Figure 1. Average-Linkage Method Dendrogram

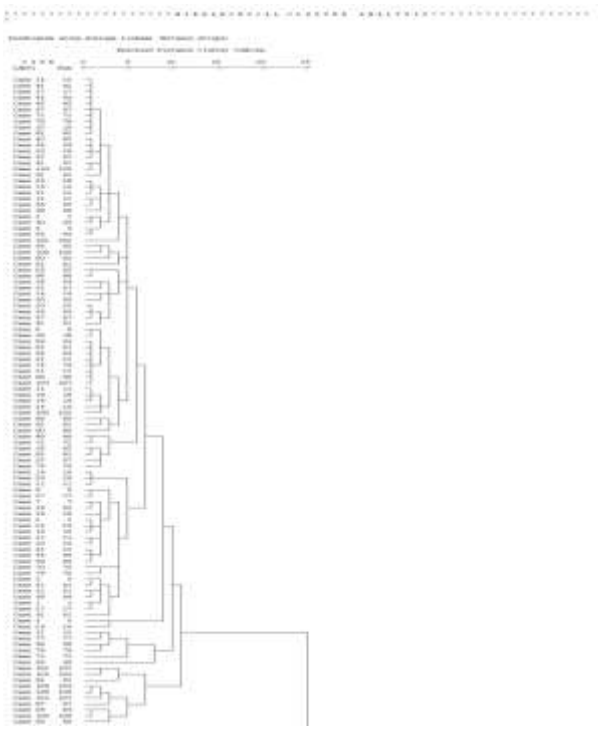
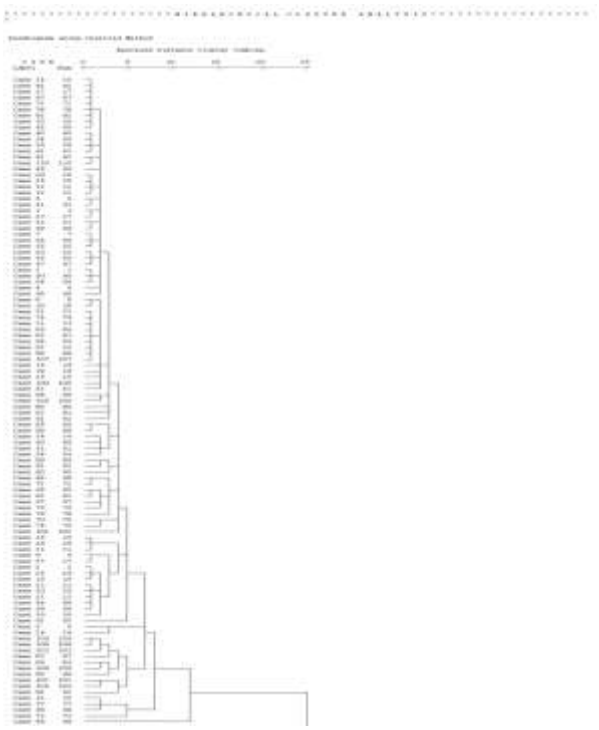
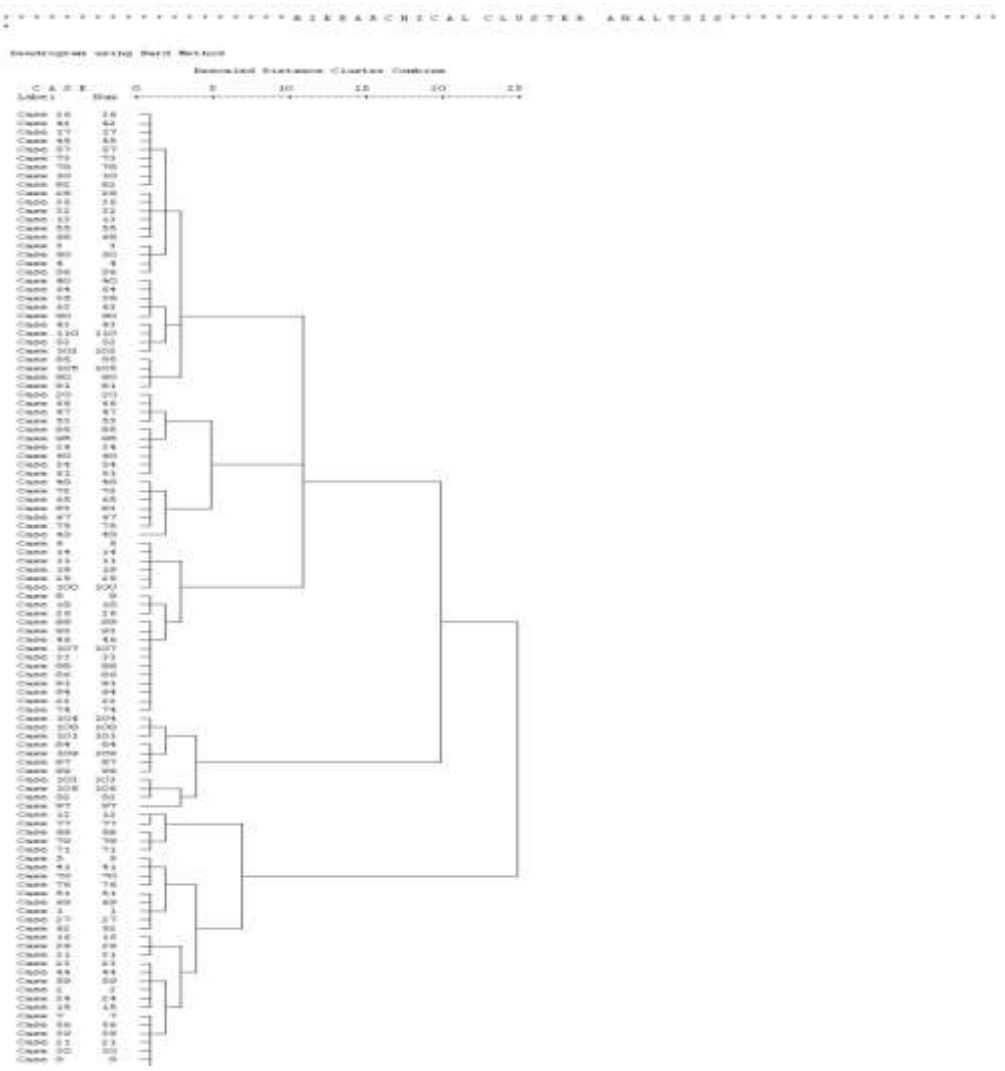


Figure 2. Centroid-Linkage Method Dendrogram



The measure used for the hierarchical cluster analysis was the interval of the squared Euclidian distance, and the values were standardized and transformed into z-scores. On the resulting Ward's linkage dendrogram, it appeared that a two- or three-cluster solution would be the most appropriate, as seen in Figure 3. Under a three-cluster solution, one cluster would have been very small, so it was decided to merge the small cluster with a larger one and continue under a two-cluster solution. A subsequent k-means cluster analysis was conducted with two clusters specified as a solution.

Figure 3. Ward's-Linkage Method Dendrogram



Following the cluster analysis, paired-sample t-tests were conducted again for each cluster, using pre- and post-test scores. The same six paired-sample variables that were used in the class-wide t-tests were again used for the cluster t-tests. Again, Holm's (1979) sequential procedure was used as a post-hoc correction method.

RESULTS

The descriptive statistics and correlations for the class-wide paired-sample t-tests can be found in Table 1. Of note, the value for the 2000-word perception pre- and post-tests represent the number of *unknown* words, hence the decreasing number in the post-test. Also, correlations between the paired-sample variables were strong as expected, since each paired-sample tested the same construct through a pre- and post-test. The N-size for the class-wide paired-sample t-tests was 110.

Table 1. Descriptive Statistics and Correlations for Class-Wide Paired-Sample T-Tests

		Mean	*SD*	*SE*	*Corr.*	*Sig.*
Pair 1	2000-word perception post	20.84	22.54	2.15	.51	.000**
	2000-word perception pre	137.34	86.79	8.28		
Pair 2	2000-word production post	10.75	3.26	.31	.39	.000**
	2000-word production pre	10.77	3.21	.31		
Pair 3	2000-word reception post	24.65	2.79	.27	.61	.000**
	2000-word reception pre	24.70	2.66	.25		
Pair 4	Reading ability post	17.17	2.67	.25	.41	.000**
	Reading ability pre	14.57	2.62	.25		
Pair 5	English ability post	15.94	3.27	.31	.42	.000**
	English ability pre	12.34	3.14	.30		
Pair 6	Words-per-minutes post	128.28	33.73	3.22	.54	.000**
	Words-per-minute pre	119.09	35.36	3.37		

*** significant using Holm's (1979) sequential procedure (beginning at 0.05)*

The results of the paired-sample t-tests for all of the participants can be found in Table 2. On average, participants experienced significantly higher scores on the post-tests for the three affective measures (Pair 1, Pair 4, Pair 5) when compared to their pre-test scores. Also of note, the effect size for these three pairs was considerable, based on the guidelines suggested by Tabachnick and Fidell (2007), for which a small r^2 is around .01, a medium-effect size is around .09, and a large-effect size is around .25. The squared-effect sizes for pairs 1, 4, and 5 were .48, .20, and .27, respectively. With regard to reading fluency, participants experienced a significant increase on the post-test (Pair 6). However, it should be noted that the effect size for this increase in reading fluency was rather small. Finally, participants did not experience any significant increases in terms of actual vocabulary knowledge, as noted on the 2000-word production and reception post-tests (Pair 2, Pair 3).

After conducting class-wide paired-sample t-tests, a cluster analysis was performed in order to segment the class into different student clusters. The resulting dendrogram (Figure 3) indicated that a two-cluster solution was the most appropriate for the available data. Results of the cluster analysis are summarized in Table 3a. Co-linearity diagnostic tests were also conducted, reported in Table 3b, to ensure that all of the variables had acceptable tolerance levels above 0.2 (Field, 2005). The larger of the two clusters, cluster one (with an N-size of 75 students), scored higher on all of the variables. Cluster one had better vocabulary knowledge, higher perceived reading and English ability, and read more words-per-minute. The only area in which cluster one had a lower score than cluster two was in the 2000-word level perception vocabulary test, which again indicates a superior perception of ability because this score indicates *unknown* words. As a result, a good way of characterizing these students may be as "higher-level" (cluster one) and "lower-level" (cluster two).

The descriptive statistics and correlations for cluster one's paired-sample t-tests can be found in Table 4. Of note, correlations between the paired-sample variables were generally weaker for cluster one than they were for the class-wide paired-sample t-tests. The N-size for cluster one's paired-sample t-tests was 75.

Table 2. Class-Wide Paired-Sample T-Tests

		Mean	*SD*	*SE*	*95% CI* *lower*	Upper	*T*	*df*	*Sig.*	*r*
Pair 1	2000-word perception post 2000-word perception pre	-116.50	77.82	7.42	-101.79	-131.21	-15.70	109	.000**	.69
Pair 2	2000-word production post 2000-word production pre	-.02	3.55	.34	.65	-.69	-.05	109	.957	.00
Pair 3	2000-word reception post 2000-word reception pre	-.05	2.40	.23	.41	-.50	-.20	109	.843	.00
Pair 4	Reading ability post Reading ability pre	2.60	2.86	.27	3.14	2.06	9.52	109	.000**	.45
Pair 5	English ability post English ability pre	3.60	3.45	.33	4.25	2.95	10.96	109	.000**	.52
Pair 6	Words-per-minutes post Words-per-minute pre	9.19	33.12	3.16	15.45	2.93	2.91	109	.004**	.07

*** significant using Holm's (1979) sequential procedure (beginning at 0.05)*

Table 3a. Cluster Means

	Cluster		*Cluster Z-scores*	
	1 (75N)	*2 (35N)*	*1 (75N)*	*2 (35N)*
2000-word perception pre	87.43	244.29	-.58	1.23
2000-word production pre	11.20	9.83	.14	-.29
2000-word reception pre	25.36	23.29	.25	-.53
Reading ability pre	14.75	14.20	.07	-.14
English ability pre	12.49	12.00	.05	-.11
Words-per-minute pre	123.71	109.20	.13	-.28
Total pages read	393.96	405.06	-.04	.08

Table 3b. Co-linearity Diagnostics

	Co-linearity statistics	
	Tolerance	*VIF*
2000-word perception pre	.85	1.18
2000-word production pre	.67	1.50
2000-word reception pre	.67	1.49
Reading ability pre	.62	1.62
English ability pre	.65	1.55
Words-per-minute pre	.84	1.20

Table 4. Descriptive Statistics and Correlations for Cluster One's Paired-Sample T-Tests

		Mean	*SD*	*SE*	*Corr.*	*Sig.*
Pair 1	2000-word perception post	14.39	15.19	1.75	.46	.000**
	2000-word perception pre	87.43	44.00	5.08		
Pair 2	2000-word production post	11.00	3.26	.38	.28	.014**
	2000-word production pre	11.20	2.95	.34		
Pair 3	2000-word reception post	25.14	1.88	.22	.23	.047
	2000-word reception pre	25.36	1.50	.17		
Pair 4	Reading ability post	17.22	2.58	.30	.37	.001**
	Reading ability pre	14.75	2.53	.29		
Pair 5	English ability post	16.25	3.38	.39	.43	.000**
	English ability pre	12.49	3.15	.36		
Pair 6	Words-per-minutes post	131.57	33.56	3.88	.51	.000**
	Words-per-minute pre	123.71	37.76	4.36		

*** significant using Holm's (1979) sequential procedure (beginning at 0.05)*

The results of the paired-sample t-tests for cluster one can be found in Table 5. Again, affective measures (Pair 1, Pair 4, and Pair 5) were significantly higher on the post-tests than on the pre-tests. Effect sizes were also considerable. Of note, reading fluency (Pair 6) did not achieve significance, and after using the Holm's (1979) sequential procedure post-hoc test, it was not as close to significance as it initially appeared (with the significance threshold for Pair 6 settling at 0.017, well below the actual 0.058 indicated). Finally, cluster one participants did not experience any significant increases in terms of actual vocabulary knowledge, as noted on the 2000-word production and reception post-tests.

Table 5. Cluster One Paired-Sample T-Tests

		Mean	*SD*	*SE*	*95% CI* *lower*	*upper*	*T*	*df*	*Sig.*	*r*
Pair 1	2000-word perception post 2000-word perception pre	-73.04	39.43	4.55	-63.97	-82.11	-16.04	74	.000**	.77
Pair 2	2000-word production post 2000-word production pre	-.20	3.73	.43	.66	-1.06	-.47	74	.644	.00
Pair 3	2000-word reception post 2000-word reception pre	-.21	2.12	.24	.27	-.70	-.87	74	.386	.10
Pair 4	Reading ability post Reading ability pre	2.48	2.87	.33	3.14	1.82	7.49	74	.000**	.43
Pair 5	English ability post English ability pre	3.76	3.50	.40	4.57	2.95	9.30	74	.000**	.54
Pair 6	Words-per-minutes post Words-per-minute pre	7.87	35.40	4.09	16.01	-.28	1.93	74	.058	.05

*** significant using Holm's (1979) sequential procedure (beginning at 0.05)*

The descriptive statistics and correlations for cluster two's paired-sample t-tests can be found in Table 6. Of note, correlations between the paired-sample variables were stronger than for both cluster one and the class-wide sample, with the exception of pairs 1 and 5. Also, the first pair did not achieve a significant correlation (Pair 1). The N-size for cluster two's paired-sample t-tests was 35.

Table 6. Descriptive Statistics and Correlations for Cluster Two's Paired-Sample T-Tests

		Mean	*á.*	*SE*	*Corr.*	*Sig.*
Pair 1	2000-word perception post 2000-word perception pre	34.66 244.29	28.95 51.64	4.89 8.73	.19	.274
Pair 2	2000-word production post 2000-word production pre	10.20 9.83	3.24 3.56	.55 .60	.58	.000**
Pair 3	2000-word reception post 2000-word reception pre	23.60 23.29	3.95 3.84	.67 .65	.72	.000**
Pair 4	Reading ability post Reading ability pre	17.06 14.20	2.89 2.81	.49 .47	.49	.003**
Pair 5	English ability post English ability pre	15.26 12.00	2.96 3.15	.50 .53	.40	.016**
Pair 6	Words-per-minutes post Words-per-minute pre	121.23 109.20	33.46 27.51	5.66 4.65	.60	.000**

*** significant using Holm's (1979) sequential procedure (beginning at 0.05)*

The results of the paired-sample t-tests for cluster two can be found in Table 7. Again, affective measures (Pair 1, Pair 4, and Pair 5) were significantly higher on the post-tests than on the pre-tests. Effect sizes were also considerable. Also of note, reading fluency (Pair 6) achieved significance (with the Holm's (1979) procedure threshold settling at 0.017). This is a key divergence from cluster one, which did not achieve significance in terms of reading fluency. Finally, cluster two participants did not experience any significant increases in terms of actual vocabulary knowledge, as noted on the 2000-word production and reception post-tests.

Table 7. Cluster Two Paired-Sample T-Tests

		Mean	*SD*	*SE*	*95% CI* *Lower*	Upper	*T*	*df*	*Sig.*	*r*
Pair 1	2000-word perception post 2000-word perception pre	-209.63	54.19	9.16	-191.01	-228.24	-22.89	34	.000**	.94
Pair 2	2000-word production post 2000-word production pre	.37	3.13	.53	1.45	-.71	.70	34	.488	.01
Pair 3	2000-word reception post 2000-word reception pre	.31	2.92	.49	1.32	-.69	.64	34	.528	.01
Pair 4	Reading ability post Reading ability pre	2.86	2.88	.49	3.85	1.87	5.87	34	.000**	.50
Pair 5	English ability post English ability pre	3.26	3.35	.57	4.41	2.10	5.76	34	.000**	.49
Pair 6	Words-per-minutes post Words-per-minute pre	12,03	27.88	4.71	21.61	2.45	2.55	34	.015**	.16

*** significant using Holm's (1979) sequential procedure (beginning at 0.05)*

The divergence in reading fluency between cluster one and cluster two is illustrated in Figure 4 and Table 8. Reading fluency was assessed at three points during the study. The first and final points in Figure 4 are the same as the information included in Tables 4 and 6. In addition, a fluency score was assessed in the middle of the semester, for which cluster one had a mean average of 120.67 and cluster two had a mean average of 110.20. It should also be noted that when a one-way ANOVA was conducted between the clusters with regard to reading fluency, indicated in Table 8, there was a significant difference between the clusters when measured at the beginning of the semester, $F\ (1, 108) = 4.13, p < 0.05$. However, there was not a significant difference between the clusters when measured in the middle of the semester $F\ (1, 108) = 2.76, p > 0.05$ or at the end of the semester, $F\ (1, 108) = 2.27, p > 0.05$.

Figure 4. Reading Fluency Progress for Both Clusters

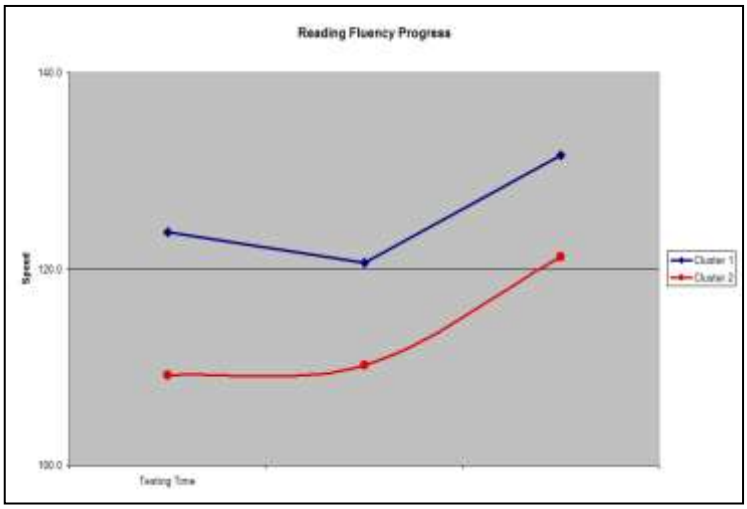


Table 8. One-Way ANOVA on Reading Fluency Measures

		Sum of Squares	Df	Mean Square	F	Sig.
Words-per-minute pre	Between Groups	5021.94	1	5021.94	4.13	.05
	Within Groups	131261.15	108	1215.38		
	Total	136283.10	109			
Words-per-minute mid	Between Groups	2614.29	1	2614.29	2.76	.10
	Within Groups	102280.27	108	947.04		
	Total	104894.56	109			
Words-per-minute post	Between Groups	2553.75	1	2553.75	2.27	.14
	Within Groups	121422.52	108	1124.28		
	Total	123976.26	109			

DISCUSSION

The first hypothesis predicted significant gains in student affect and reading fluency, as measured by a vocabulary perception test, a reading ability perception questionnaire, an overall English ability perception questionnaire, and a WPM fluency test. As predicted, the class-wide sample saw significant gains in both areas (see Tables 1 and 2).

Also as predicted, vocabulary gains measured by Nation's 2000-word level production and reception tests did not significantly improve. The nature of ER, diverging from more traditional methods of instruction, seemed to have a quick and profound impact upon student affect. After six years of grammar-translation exam preparation in junior- and senior-high school, it is not surprising that students' affect improved so dramatically, considering the emphasis ER places on student enjoyment. However, it would be interesting to see if these gains could be sustained over a longer timeline. It is likely that to permanently change student perceptions of their vocabulary, reading, and English abilities, they would need to experience more than a single semester of ER. If delayed post-tests were conducted a few months after the end of the semester, it is doubtful that affect scores would remain as high. The dramatic increase in affect over the semester suggests that perception is a relatively unstable variable, and to generate permanent change, it would likely take years of nurturing students' confidence.

With regard to fluency, there was a significant increase, yet the effect size of .005 was disappointingly small. Again, a longer study may yield additional insights as to the rate of fluency improvement. It would seem that fluency should eventually level-off as students approach the upper-limits of their capabilities. However, in the case of this study their fluency actually appeared to accelerate as they approached the end of the semester, with a pre-semester reading speed of 119 WPM (as seen in Table 1 and Figure 4), a mid-semester reading speed of 117 WPM (as seen in Figure 4), and a post-semester reading speed of 128 WPM (as seen in Table 1 and Figure 4). Research has suggested that the fluency upper limits for university students studying language, albeit ESL, is in the neighborhood of 200 WPM (Nuttall, 1982), which would discount the suggestion that cluster one had reached their ceiling.

It is also worth noting that the multiple choice comprehension tests, used to ensure that readers read properly rather than speed-read, may have influenced this study's findings, even though they were not used as a variable in the study. Particularly, after students completed the first multiple-choice comprehension test at the beginning of the semester, they may have modified their reading style during subsequent reading assessments to better fit the requirements of the multiple-choice testing method. For instance, in the second and third reading assessments, instead of reading thoroughly, students may have engaged in more skimming, paying attention to specific points most likely to appear on a multiple-choice comprehension test (such as names, places, numbers, and descriptions). There is a reasonable possibility that fluency scores at the

end of the semester for both clusters may have been enhanced by this increased familiarity with the testing method. However, it is less clear whether one of the clusters had a greater advantage because of this increased testing familiarity. It is therefore conceivable that the higher cluster possessed superior study strategies, and was better able to adapt their reading style during subsequent reading assessments. In such a case, the actual gain shown by the lower cluster in reading fluency would have been even more pronounced if the strategy advantage held by the higher cluster was considered. Conversely, it is also conceivable that the lower cluster was more motivated to find alternative strategies to compensate for their lesser degree of comprehension, in which case, the significant finding in this study would have been nullified. Either way, it is an interesting question worth exploring in future research.

A final point to address for the fluency measure is the dip in reading speed during the middle of the semester for the higher cluster, while the lower cluster maintained a similar fluency score as their pre-test. One plausible explanation could be that the higher cluster was more prone to adjusting reading speed in relation to comprehension, because of a greater awareness of their learning capabilities. Since the mid-semester reading was more difficult, evidenced in that the lexical coverage of the text did not satisfy Hu and Nation's (2000) suggested 95% coverage that would ensure fluent reading, the higher cluster may have realized that they had to read at a slower pace in order to comprehend the passage, and may have been quicker to adjust their behavior accordingly. Conversely, the lower cluster may not have been as aware of the reading comprehension and hence failed to adjust their reading speed. A closer examination of comprehension awareness and the subsequent modification of behavior is certainly something worth exploring in future research.

Finally, claims in previously-mentioned research of vocabulary knowledge not improving with ER appear to be valid. In fact, vocabulary knowledge scores actually *decreased* over the semester, although this was not statistically significant. There may have been a couple of reasons behind this lack of vocabulary development. It is possible that the emphasis on communicative activities during class time reduced the amount of time students could have spent reading. During the study, it was assumed that the communicative activities may have complemented the reading by providing multi-disciplinary opportunities for recycling. However, it is certainly possible that students were not actually recycling the vocabulary they read, and were missing opportunities at recycling through more reading. Another possible reason for this lack of vocabulary gain may have been because students were free to choose graded readers from various publishers, specifically Penguin and Oxford. It is equally possible that each publisher emphasized different reading lists, resulting in less repetition of key vocabulary. If students had been limited to a single publisher, there may have been more vocabulary recycling and, as a result, greater vocabulary acquisition. This, also, is certainly worth exploring in future research.

Taking all three variables into account, Tables 1 and 2 seem to suggest a benefits-hierarchy, with increased student affect as the quickest and most pronounced byproduct of ER, followed by slower and smaller increases in reading fluency, and vocabulary knowledge not improving at all. It is important to remember that these are the results of a communicative teaching approach with reading done outside of class. It was hoped that the use of random interviews to confirm student reading done outside of class, coupled with reassurance from the teacher that infrequent reading would not harm students' grades, produced relatively accurate reported reading results. An alternative teaching methodology may yield different results, and would, as well, certainly be a worthy course of study for future researchers.

These findings have pedagogical implications in that single one-off ER courses, offered by many universities in Japan, are likely not enough to help boost students' linguistic abilities. Unfortunately, educational institutions that suggest students enroll in ER classes to assist in preparation for proficiency tests (like TOEFL) may be better served by offering intensive reading courses, or at the very least, coordinating ER courses with complementary English courses,

ensuring the recycling of concepts, grammar, and vocabulary. In this particular instance, ER appeared to do little more than enhance students' perceptions of their abilities, which may be a worthy goal in some situations. However, it is important to remember that this sample came from a single university, which may have skewed the results. The participants involved may have shared similar characteristics such as motivation, anxiety, conscientiousness, or other individual differences, which resulted in their admission into the same university. If this study was repeated with students from a more demanding university, students may have been more conscientious and may have reaped more linguistic benefits from ER through taking notes, reviewing what they read, or using a dictionary. Conversely, involving a school with less conscientious students (or other individual differences) may have equally resulted in the opposite scenario.

The cluster analysis essentially divided the class into two-thirds and one-third groups, with standardized z-scores of six variables creating a "higher-level" cluster of 75 and a "lower-level" cluster of 35 (Table 3a). Initial differences between the clusters were most pronounced in terms of linguistic abilities (i.e., productive vocabulary, receptive vocabulary, and reading fluency), and not as pronounced in affective areas (particularly in their perceptions of their reading and English abilities). Both clusters developed in a similar fashion over the course of the semester, except with regard to fluency. During the study, the "lower-level" cluster narrowed the reading fluency gap with the "higher-level" cluster, suggesting that perhaps ER may yield more benefits for lower-level students. To confirm this, additional research would need to be conducted, possibly involving students from different class levels in order to accentuate their differences. In this study, the two clusters began the semester as significantly different groups, but ended the semester as significantly similar. The higher-level cluster's gain did come close to an uncorrected significance threshold of 0.05, but after the post-hoc correction brought the significance level down to 0.017, they were not as close as initially thought. Conversely, the lower-level cluster just barely achieved statistical significance, coupled with a larger effect size.

While it is not entirely clear as to why the lower-level students' fluency increased at a greater rate, there are nonetheless a few possibilities that could be explored in future research. One possibility is that the wide-spread emphasis on creating interest in ER for lower-level students may have resulted in an accidental neglect of higher-level students. With publishers increasingly aiming for students at the lower end of the spectrum, with a larger selection of very easy graded readers, it is possible that the availability of more challenging graded readers has not been able to keep pace. It would be interesting to examine publisher title-lists and see exactly how many titles are available at each reading level. Further, there seems to be a great deal of emphasis on making the easier titles more accessible, through more pictures and more vivid color pagination. Again, it would be interesting to see if more challenging graded readers have been able to keep up with the aesthetic enhancements made to easier graded readers. Future research could examine if there are correlations between available graded reader titles at each level, aesthetic enhancements of graded readers, reading fluency, total pages read, level of challenge, pleasure, and effort put forth by students.

Another possible explanation for the difference in reading fluency gains between the clusters may have been that the higher-level cluster was less convinced of ER's painless approach to reading. Especially in Japan, a country where students endure long years of arduous study, the higher-levels may have been more likely to subscribe to a *no-pain, no-gain* reading philosophy. While lower-levels may have been consumed with the success they were finally starting to feel in an English class, the higher-levels may have felt that the reading was far too easy for them, especially since they had been successful earlier in their scholastic careers with far more demanding tasks. The lower-level cluster actually read more pages and had a larger increase in reading ability perception than the higher-level cluster, although neither of these differences was large enough to register statistical significance.

One final explanation may be that reading fluency eventually begins to level-off as students reach the ceiling of their abilities, giving an advantage to students farther away from the ceiling. However, with the higher-level cluster only reading at 131 WPM, it is debatable as to whether they were approaching the ceiling of their abilities. Another possibility, as pointed out earlier, may be that the higher-level students were more conscious of the multiple-choice comprehension quiz they would have to do after the reading, perhaps because of superior meta-awareness, and consequently slowed their speed, resulting in a muted WPM score.

Concerns and Limitations

It is important to note that cluster analysis is often used as an exploratory statistical procedure that may exaggerate differences in order to create clusters (Kettenring, 2006). While cluster analysis can provide a number of valuable insights, and the number of published articles using cluster analysis has increased a great deal recently, it is important to follow-up cluster analysis with a statistical analysis that is more conservative in its stance (Kettenring, 2006). In the case of this study, the paired-sample t-tests provided a more conservative basis for comparison, as opposed to the cluster means generated in the cluster analysis. Also, it should be noted that outside influences may have had an impact upon the results obtained here. Participants were controlled within the institution, but any number of factors may have influenced the results from outside the research institution (for example, private tutoring, enrollment at another institution, or time spent abroad during the study). Nevertheless, we feel that this research did provide interesting insights with significant pedagogical value.

The results of this research showed a significant class-wide improvement in reading fluency, but the abbreviated duration of the study likely truncated the reading fluency improvement. A longer study, perhaps over an entire year (two semesters or more), may have generated more pronounced reading fluency increases, and may have resulted in an even larger gap reduction between the two clusters. Additionally, vocabulary knowledge increases may have become evident over a longer research period, especially since their acquisition is contingent upon recycling. Finally, a longer study with delayed post-tests may have revealed the permanence or impermanence of the student affect increase. As the most positive result in this study, it is crucially important to determine if these affective gains will last.

Further, in order to maintain a healthy sample size for the second part of the analysis involving cluster comparisons, it was thought that introducing a control group for the first class-wide analysis should not be pursued. However, comparing the effects of different ER approaches on student gains in affect, vocabulary, and reading fluency is under-explored and very important. Future research may want to replicate this study, but with a larger sample size that allows for multiple conditions, such as only reading with no supplementary in-class activities or, if activities are desired, perhaps less rigid ones, such as simple free writing exercises, which would ensure that students actually complete the reading. Additionally, a qualitative element to this study may yield additional insights, such as how ER made students feel.

There were also some problematic testing issues involving the sensitivity of the vocabulary tests and the frequency of the fluency tests. Nation's 2000-word level productive and receptive tests may have been too broad to measure the small number of new words introduced via graded readers. It is possible that students did, in fact, learn some new vocabulary which was not represented on the 2000-word level tests. A more sensitive vocabulary test that isolates key vocabulary targets within the graded readers may yield more positive results; however, any subsequent results would need to be considered in relation to the sensitivity of the vocabulary test. Unfortunately, getting access to the word-lists used by publishers is often a closely guarded industry secret, which might make such a study difficult to construct. Additionally, multiple fluency tests spread more frequently throughout the study, and averaged to create a mean score,

may have yielded more reliable fluency scores. Only testing students once at the beginning, middle, and end of the semester, allows for the possibility of an anomalous bad test that could skew the results. This may have been the reason behind the higher-level cluster's mid-semester decrease in fluency scores, as shown in Figure 4.

CONCLUSION

ER has been gaining credibility in Japan as an effective way of boosting student affect, strengthening vocabulary knowledge, and increasing reading fluency. The increasing number of ER studies based in Japan is evidence of its growing acceptance as a legitimate pedagogical approach. What the research community has not yet addressed, however, is how different ER approaches yield benefits in varying degrees, and how different students benefit in different ways. What this study has attempted to demonstrate is that gains among affect, vocabulary, and fluency are very different when examined within a single teaching framework, and that not all students follow the same learning trajectory in ER classes. By placing these issues within a practical framework, this study attempted to merge some of the established ER theory with practical pedagogical goals. At the very least, hopefully this study will prompt others to challenge these assumptions and provide additional insights as to how ER works in a practical and multi-contextual classroom setting.

Omar Karlin *is a Professor in Meiji University's English Track program where he teaches English content courses as part of the Japanese government's Global 30 initiative. His research interests are centered on how the personality of students can affect their acquisition of English, particularly when studying abroad. He is currently a doctoral candidate at Temple University Japan, and writing his thesis on the development of a language-sensitive personality test, and how personality affected the acquisition of English for a group of Japanese university students after a month of studying in Australia.*

Email: o_karlin@yahoo.com

Rick Romanko *is an Assistant Professor at Wayo Women's University in Chiba, Japan. He holds a M.Ed. in TESOL from Temple University. His research interests include developing extensive reading programs, corpora-informed vocabulary and language learning, and how learning is enhanced by various teaching methodologies and materials development.*

Email: rromanko@gmail.com

REFERENCES

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Carrell, P. L., & Carson, J. G. (1997). Extensive and intensive reading in an EAP setting. *English for Specific Purposes, 16*, 47-60.

Cobb, T. (1994). Web Vocabprofile [accessed June, 5, 2008 from http://www.lextutor.ca/vp/], an adaptation of Heatley & Nation's (1994) Range.

Day, R., & Bamford, J. (2002). Top ten principles for teaching extensive reading [Electronic Version]. *Reading in a Foreign Language*, *14*(2). Retrieved May 3, 2006, from http://nflrc.hawaii.edu/rfl/October2002/day/day.html.

Dornyei, Z. (2005). *The Psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Escott, J. (2002). *The wild west, Oxford Dominoes Reading Series*. Oxford University Press.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.

Fitzgerald, D. (2002). *True heroes of sport, Oxford Dominoes Reading Series*. Oxford University Press.

Gorusch, G. (2001). Japanese EFL teachers' perceptions of communicative, audio-lingual and yakudoku activities: The plan versus the reality. *Education Policy Analysis Archives, 9*(10), 33.

Hashimoto, M., Takada, T., Isobe, T., Sakai, N., Ikemura, D., & Yokogawa, H. (1998). Reading shido 12 no approach [Twelve approaches to teaching reading]. *Eigokyoiku [English Education]*, *47*(2), 42-43.

Heatley, A., & Nation, P. (1996). *Range* [Computer software]. Wellington, New Zealand: Victoria University of Wellington. (Available from http://www.vuw.ac.nz/lals/)

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal Statistics, 6*, 65-70.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*(1), 23-59.

Iwahori, Y. (2008). Developing reading fluency: A study of extensive reading in EFL. *Reading in a Foreign Language, 20*(1), 70-91.

Kettenring, J. R. (2006). The practice of cluster analysis. *Journal of Classification, 23*, 3-30.

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly, 13*(2), 123-147.

Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System, 25*(1), 91-102.

Mori, S. (2004). Significant motivational predictors of the amount of reading by EFL learners in Japan. *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia, 35*(1), 63-81.

Nation, P. (1990). *Teaching and learning vocabulary*. Massachusetts: Newbury House.

Nation, P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press. Nation, P., & Laufer, B. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Nuttall, C. (1982). *Teaching reading skills in a foreign language*. London: Heinemann Educational Books.

Nuttall, C. (1996). *Teaching reading skills in a foreign language* (New ed.). Portsmouth, NH: Heinemann.

Ono, L., Day, R., & Harsch, K. (2004). Tips for reading extensively [Electronic Version]. *English Teaching Forum Online*, *42*(4). Retrieved May, 3, 2006 from http://exchanges.state.gov/englishteaching/forum/archives/2004/04-42-4.html

Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, *5*(2), 239-251.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, *16*(2), 70-96.

Takase, A. (2003). *Effects of eliminating some de-motivating factors in reading English extensively.* Paper presented at the 29th Annual Japan Association for Language Teaching National Conference 2003, Shizuoka.

Takase, A. (2004). The effects of extensive reading on the motivation of Japanese high school students. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *54*(7-A), 24-29.

Takase, A. (2006). *Teachers motivated by students' extensive reading: A case study of teachers' motivation to start reading English books.* Paper presented at the 31st Annual Japan Association for Language Teaching National Conference 2005, Shizuoka.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130-163.

West, M. (1953). *A general service list of English words*. London: Longman.

Appendix A

Your Opinions about Your Reading-English Ability

Name: _____ Student Number: _____

Please be honest and answer the questions as best as you can. For each question circle one of the choices.

						
I don't think so at all まったくそう思わない **1**		**I don't think so** そう思わない **2**		**I think so** そう思う **3**		**I really think so** とてもそう思う **4**

		1	2	3	4
1	I enjoy reading in my native language.				
2	I enjoy reading in English.				
3	I am interested in reading in English.				
4	I am confident when I read in English.				
5	I can read fast in English.				
6	I am interested in reading English books for pleasure.				
7	I know many words in English.				
8	I can write well in English.				
9	I can spell well in English.				
10	I am good at English grammar.				
11	I can speak English well.				
12	I can talk (in English) about books I read.				
13	I am good at English.				

Appendix B

Reading Fluency (Form A) — Native Americans

Native Americans- or 'Red Indians' –lived in North America for thousands of years before the people from Europe arrived. There were many different tribes, most with different languages. Tribes had their lands for hunting, and sometimes they fought about them. But the land, the mountains, the rivers, and the trees were living things for them. The Indians understood the land. They took from it only the things they needed to live.

But to the 'white man', land was something to buy and sell and to make money with. White men wanted to get land for farms and to make mines, where they could look for gold under the land.

The United States government did not understand the Native Americans' love of the land-its mountains and rivers. They asked the Indians to give some of their lands to the white men, and at first they did. But then the white men wanted more and more land, and the government moved the Indians onto lands that the white men didn't want. They called these lands reservations.

Between 1853 and 1874, the Indians lost 700 million square kilometers of land to the white men. There were many wild animals on these lands. The Native Americans hunted and killed only the animals they needed, and no more. For many tribes, the most important animal was the buffalo. They ate buffalo meat, and buffalo skin became trousers, dresses and shoes for them to wear, and homes for them to live in. When the Native Americans saw the white men hunting and killing thousands of buffaloes, they were afraid and angry. Some wanted to fight with the white men.

Time from the blackboard: *273 words*

Reading Fluency (Form B) — Heroes of Sport: Cathy Freeman

Aborigines are black people from Australia. They lived in Australia thousands of years before white people arrived there from Europe. When the first Europeans arrived, they took Australia from the Aborigines.

Cathy Freeman in as an Aborigine athlete. She is the first Australian Aborigine to run for Australia in the Olympic Games. She was born on February 16, 1973, in Queensland, Australia. She was one of the fastest children at her school. When she ran, she always dreamed of being in the Olympic Games and winning a gold medal.

Cathy's family didn't have much money when she was young, and they couldn't buy shoes for her to run in. Soon Cathy became famous. People called her 'the little Aborigine girl with no shoes'. When she was eight years old she ran for Queensland in an important race for young athletes. Before the race began everyone laughed at Cathy with no shoes on her feet. But in the end Cathy won. After the race she said, 'you don't need shoes to win a race.'

Cathy won her first gold medal in 1994 at the Commonwealth Games in New Zealand. After the race Cathy wore the Aborigine flag and not the Australian one for all of her photos. Many Australian people were angry about this. They said, 'Cathy must wear the Australian flag when she runs for Australia.

But Cathy was a very good athlete and she went on running for Australia. At the 2000 Olympic Games in Sydney, she won a gold medal in the 400 meters. This time she wore the Australian flag *and* the Aborigine flag for her photos after the race!

She says, 'You don't need money to win. To win, you must go after your dreams.'

Time from the blackboard: *289 words*

Reading Fluency (Form B) — Heroes of Sport: Cathy Freeman

Aborigines are black people from Australia. They lived in Australia thousands of years before white people arrived there from Europe. When the first Europeans arrived, they took Australia from the Aborigines.

Cathy Freeman in as an Aborigine athlete. She is the first Australian Aborigine to run for Australia in the Olympic Games. She was born on February 16, 1973, in Queensland, Australia. She was one of the fastest children at her school. When she ran, she always dreamed of being in the Olympic Games and winning a gold medal.

Cathy's family didn't have much money when she was young, and they couldn't buy shoes for her to run in. Soon Cathy became famous. People called her 'the little Aborigine girl with no shoes'. When she was eight years old she ran for Queensland in an important race for young athletes. Before the race began everyone laughed at Cathy with no shoes on her feet. But in the end Cathy won. After the race she said, 'you don't need shoes to win a race.'

Cathy won her first gold medal in 1994 at the Commonwealth Games in New Zealand. After the race Cathy wore the Aborigine flag and not the Australian one for all of her photos. Many Australian people were angry about this. They said, 'Cathy must wear the Australian flag when she runs for Australia.

But Cathy was a very good athlete and she went on running for Australia. At the 2000 Olympic Games in Sydney, she won a gold medal in the 400 meters. This time she wore the Australian flag *and* the Aborigine flag for her photos after the race!

She says, 'You don't need money to win. To win, you must go after your dreams.'

Time from the blackboard: *289 words*

Reading Fluency (Mid-semester) — The Paralympic Games

The Olympic Games began in Greece about three thousand years ago. Only men could be Olympic athletes, and they wore nothing when they ran in Olympic races. The discus and the pentathlon began in these early Olympic Games. (The marathon began when Greece was at war, and a soldier ran about forty-two kilometers from a town called Marathon to Athens to tell the people there about the Greeks winning the war. The soldier died soon after he arrived in Athens.) In 394 the Romans stopped the Greek Olympic Games, because they didn't like them.

In 1896, a Frenchman –Pierre de Coubertin- began the Olympic Games again. These days the Olympic Games usually happen every four years.

The Paralympic games are Olympic Games for disabled athletes. Disabled athletes are people who cannot use some part of their body easily or completely because they have an injury or illness. The Paralympic games started after World War II. In the 1940s Sir Ludwig Guttmann was a doctor at a hospital in England. At the hospital there were many disabled soldiers from World War II and Guttmann wanted these soldiers to get better by doing sports.

In July 1948, when the Olympic Games happened in London, Guttmann asked disabled English soldiers to go to a sports meeting together at Stoke Mandeville. It was all very successful, so he did it again four years later in 1952. This time disabled soldiers from Holland came too. Because he worked a lot with disabled athletes in the 1940s and 1950s, people often call Sir Ludwig Guttmann 'the father of the Paralympics'.

The first true Paralympic Games happened in Rome in 1960. Four hundred disabled athletes from twenty-three different countries came to these games. At the Sydney Paralympics in 2000, there were 4,000 disabled athletes from 122 countries!

Time in seconds: *303 words*

Comprehension Questions (Form A)

Name_____ Student Number _____
Male Female

Answer the following question about the passage you just read. Choose the best answer.

Why did the white men want to get land from the Indians?

A. For hunting and food.
B. For farming and mining.
C. For daily living.

What did the government do when the white men wanted more and more land?

A. They asked the Indians for more land.
B. They fought with the Indians and took their land.
C. They moved them onto reservations.

How much land did the Indians loose to the white men?

A. 874 million square kilometers.
B. 700 million square kilometers.
C. 853 million square kilometers.

What did the Indians use buffaloes for?

A. Hunting and eating.
B. Everything in their daily lives.
C. Clothes and shoes.

Why were the Indians angry and afraid when they saw the white men hunting and killing thousands of buffaloes?

A. Because the Indians liked hunting the buffaloes.
B. Because the Indians needed the buffaloes to live.
C. Because the Indians thought the white men would kill them next.

(X ÷ Y x 6 = WPM) **Words per minute: _____** **Score on the questions: _____**

Comprehension Questions (Form B)

Name_____ Student Number _____
Male Female

Answer the following question about the passage you just read. Choose the best answer.

Who are the Aborigines?

A. Travelers from Europe.
B. The native people of Australia.
C. Athletes from Australia.

When did Cathy win her first gold medal?

A. 1974
B. 1994
C. 2000

Cathy Freeman thinks…

A. You need money to win a race.
B. You need to follow your dreams to win a race.
C. You need shoes to win a race.

Where did Cathy go to elementary school?

A. Sydney
B. New Zealand
C. Queensland

Why did Cathy wear the Aborigine flag after she won the race at the Commonwealth games?

A. Because she was proud to be Aborigine.
B. Because she liked the Aborigine flag better than the Australian flag.
C. Because she wanted to make Australians angry.

(X ÷ Y x 6 = WPM) **Words per minute: _____** **Score on the questions: _____**

Comprehension Questions (Mid-Semester)

Name_____ Student Number _____
Male Female

Answer the following question about the passage you just read. Choose the best answer.

The Olympic event called "the marathon" was named after

A. a person
B. a town
C. a country

Sir Ludwig Guttmann was a

A. soldier
B. doctor
C. athlete

When were the first "organized" Paralympics held?

A. 1948
B. 1952
C. 1960

Who began the Olympic Games again in the late 19th century?

A. a Roman
B. a Frenchman
C. a Greek

Who attended the first sports meeting for disabled soldiers?

A. Soldiers from Holland.
B. Soldiers from Greece.
C. Soldiers from England.

(X ÷ Y x 6 = WPM) **Words per minute: _____** **Score on the questions: _____**