Shared Identities: Our Interweaving Threads

Revisiting the extensive reading effect on TOEFL scores

David Williams
Yokohama City University

Reference data:
Williams, D. (2009). Revisiting the extensive reading effect on TOEFL scores.
In A. M. Stoke (Ed.), *JALT2008 Conference Proceedings*. Tokyo: JALT.

The expanded use of standardized testing (such as TOEFL) and the popularity of extensive reading (ER) are significant trends is Japanese tertiary education. There is a large body of evidence supporting the benefits of ER but its effects in raising the scores of such standardized tests among poorly motivated students in an L1 setting is less well understood. Using a large sample (n=295) of TOEFL scores from university level students, it was found that moderate ER can be highly effective in helping to raise TOEFL scores. It may also contribute to reducing test score attrition. In light of these findings ER is recommended as a means to help raise TOEFL scores. Gender differences in test score increases, and also between high and low achievers indicate due consideration of the student profile is needed prior to implementing a university-based ER programme.

TOEFLといった共通テストの利用拡大とエクステンシブリーディング（ER)の人気が、日本の高等教育において特筆した傾向となっている。ERの効能を裏付ける証拠は数多い、だが自国で暮らす、モチベーション度が低い学生達のテストスコアを伸ばそうとする場合に、ERが効果的であることは、さほど認識されていない。適切なERがTOEFLスコアを伸ばすのには、大変効果的であることが、大学生レベルを引き合いにしたテストスコアの数多い実例から判った。性別でテストスコアの伸びには差があり、さらに成績優秀者と非成績優秀者間にも性別差が存在する。このことは、学生が持つ背景を事前に充分考慮する必要性を、ERプログラム実施大学に対し示唆するものである。

TOEFL scores and extensive reading

The use of standardized testing tools such as TOEFL and TOEIC is a growing trend in Japanese universities. As the size of the student age population in Japan declines such tests help universities to raise their academic stature and sustain their English language programmes in the face of competition from other emerging languages and subjects. Students are also promoting TOEIC and TOEFL's popularity as the tests provide officially recognized proof of language skills and give students an advantage

in the employment market. These developments have led some institutes to adopt TOEFL as a benchmark to judge their students' language proficiency.

At the same time as standardized testing becomes more ubiquitous universities have increasingly recognized the benefits supplementary study can bring to their English programmes. This might be through the use of computer e-learning, the creation of English lunchtime clubs or self-study such as by extensive reading (ER). The lattermost, ER, has become an intrinsic part of many Japanese university curricula over the last 10 years and its beneficial effect on student motivation, vocabulary and idea development as well as overall English proficiency is already widely appreciated and understood (Day, Omura, & Hiramatsu, 1991; Elley, 1991; Krashen, 2004; Nation, 1997; Takase, 2008; Tsang, 1996; Waring & Takaki, 2003). These developing themes of both more widespread use of ER and TOEFL have given rise to an important EFL question: How well can ER translate its benefits into TOEFL scores?

Attempts to answer this question indicate that extracurricular or free (i.e. extensive) reading *are* good predictors of TOEFL scores in motivated and proficient EFL learners preparing for overseas study or in students reading in an L2 environment (Constantino, Lee, Cho, & Krashen, 1997; Gradman & Hanania, 1991). ER is also effective with enthusiastic volunteers over short time periods in an L1 environment (Mason, 2006). However, the question of how well ER might benefit TOEFL scores in a large group of *poorly* motivated students with no overseas study plans in an L1 environment remains unclear. By examining 295 students studying at a Japanese university the current research attempts to answer this and establish to what extent extensive reading can help to raise TOEFL scores.

TOEFL and extensive reading at YCU

In 2005, as part of its university-wide reorganization, Yokohama City University (YCU) implemented a new general English programme, Practical English (PE). This programme requires all students to acquire TOEFL 500 (or an equivalent TOEIC 600) to gain credit for general English and to progress to their second year studies. Students receive no teacher grades; credit is based entirely on the TOEFL score.

Students take an ITP TOEFL upon induction to the university in April and are streamed into high (above TOEFL 440) and low (TOEFL 440 and below) class groups. TOEFL is retaken at the end of the semester and students who achieve TOEFL 500 gain their English language credit as well as access to advanced level English classes and their respective majors. Those not passing at the end of the first semester in July are re-streamed for second semester PE classes and retake TOEFL at the end of the second semester in February – the university also provides opportunities to take TOEFL in September, November and March. Students continue to take PE until they achieve TOEFL 500.

The PE programme consists of three 90-minute classes per week (taught by native English and Japanese instructors), a weekly compulsory e-learning component and extensive reading. At the time of data collection (early 2007) ER books (1000 books, 200 different titles, on average 40 pages per book) were housed in a reading salon located next to the e-

learning centre. The salon was open at lunchtime and staffed by an overseas student. Students were taken to the salon as part of their class session.

Data collection and procedure

Second semester and whole year TOEFL ITP scores for all first year PE students (April 2006-February 2007) were collected together with the number of books each student read. Due to the difficulties of getting students to read (Holden, 2003; Takase, 2008), in the current research a convenience sample of 4 classes – approximately 100 students – taught by the author were asked to read at least 5 readers during the second semester. After eliminating students who had achieved TOEFL 500 or had failed to take the test in either testing period, students (n=295) were then categorized as extensive reader students (ERS, n=76) – those who read 3 or more books – or non-extensive reader students (NERS, n=219) – those who read 2 books or fewer. The sample is summarized with second semester TOEFL score changes in Table 1.

To explore the possible effect of ER, data analysis focused on the differences between students who read and those who didn't and changes in TOEFL score for the second semester and whole year. Using the non-readers as a control group, differences between males and females and between students in the high and low groups were examined. Annual data examined the number of pages of reading required to raise scores by one TOEFL point. Data was analyzed using EXCEL.

Results

Frequency of extensive reading

The frequency of extensive reading amongst ERS was relatively moderate and students read an average of 4.8 books during the second semester. This compares favorably with the first semester (4.1 books) when ER was a voluntary activity. There was little difference between the frequency of reading by males (4.8 books) and females (4.9 books) or between the high and low groups. Students reading in both semesters read an average of 9.1 books (9.3 amongst high group students). The large numbers of students who did not read (NERS, n=219), the failure by 20% of students to sit the TOEFL ITP, poor class attendance (less than 50% students attended class), and the considerable anecdotal evidence from students that TOEFL 500 was simply a means to an end indicate that students in the current research were poorly motivated.

Second semester TOEFL score results

Table 1 below shows how ERS and NERS TOEFL scores changed during the second semester. ERS raised scores by an average of 16 points (male +14pts, female +18pts) to TOEFL 454 while NERS scores *decreased* on average by -1.5 (male -3, female +0.2) to TOEFL 457.5. These score changes do not however tell us about the distribution of the scores amongst readers and non-readers.

Table 1. Summary of sample and score changes (second semester)

	NERS		ERS	
	Number of students	TOEFL score change (points)	Number of students	TOEFL score change (points)
Male	109	459⇨456 = -3	35	438.5⇨452.5 = +14
Female	110	459⇨459.2 = +0.2	41	438.5⇨456.5 = +18
Total	219	459⇨457.5 = -1.5	76	438⇨454 = +16

Score differences in second semester

Table 2 shows the distribution of second semester TOEFL score changes for ERS and NERS in increments of 20 points. Following from Table 1, we can see that ERS increased their scores more than NERS (77% against 46% respectively). For students with score gains of 20 points or more, ERS outperformed NERS by a factor greater than two (47% of ERS against 22% of NERS). However, for score increases of 1 to +19 points there was less difference between ERS and NERS (30% against 24% respectively). Concerning attrition far fewer ERS experienced score losses than NERS, indeed for those students whose scores fell by more than 20 points, NERS were almost 3 times as likely to experience such score attrition as ERS (23% against 8%).

On the basis of these observations extensive reading appears to assist in both raising scores and preventing attrition. A Chi square (X2) calculation to test this delivered an X2 statistic of 23 which is significant at the p=0.01% level, thus the hypothesis that there is no association between extensive reading and TOEFL scores should be rejected.

Second semester large score gains and falls

While moderate score gains (+1 to +19) were similar for ERS and NERS, according to Table 2 the extremes of the sample (i.e. large changes of score upwards or downwards) show considerable differences between readers and non-readers. To investigate this further and understand how ER might be associated with any such changes the standard deviation was used as a measure to discover whether large

Table 2. TOEFL score gains (n=295)

TOEFL score increase/ decrease	TOEFL score change (July – Feb)	NERS (n=219)		ERS (n=76)	
		% of NERS	Combined % NERS	% of ERS	Combined% ERS
TOEFL increase	20 or more	22%	46%	47%	77%
	1-19	24%		30%	
No change	0	8%	8%	5%	5%
TOEFL decrease	1-19	23%	46%	10%	18%
	20 or more	23%		8%	

score gains and large score attrition in the sample might be greater than normal in readers and non-readers.

Table 3 below shows the percentage of students who achieved scores either one standard deviation or more *above* the mean TOEFL score increase, or one standard deviation *below* the mean score increase. Firstly, we can see that 18% of ERS made score gains ≥+1SD (i.e. score gain ≥37). In a normally distributed sample we can expect 16% of the sample to be x >±1SD, thus we can infer here that ERS are slightly skewed towards achieving more large score gains than we would normally expect. Concerning score attrition (score losses ≥-1SD) the ERS sample is marginally skewed towards a *reduction* of score attrition (14%). Thus students who read achieved more, high scores gains and at the same time experience less high score attrition than we would expect normally.

As shown these findings are not the same for males and females however. For score differences ≥+1SD, ERS females skew towards getting considerably more higher scores than we would expect normally (22%) than males (14%). By contrast, for score *attrition* ≥-1SD, ERS females are almost normally distributed (17%) while male attrition is somewhat less than expected (11%). Thus females who read are more likely to achieve more large score gains than males, whereas in males extensive reading seems to help prevent large TOEFL score loss.

In NERS, as we might expect, both females and males skew towards fewer large score gains with the tendency stronger in males than females (11% and 14.5% respectively). In NERS score attrition ≥-1SD was distributed normally with no significant difference between males and females (16.5% and 15.5% respectively). Thus we can suggest while students who do not read are less likely to make large score gains it is not necessarily mirrored by an increase in the frequency of large score attrition.

Second semester low group and high group (LG/HG)

Table 4 describes the second semester high (TOEFL ≥441) and low (TOEFL ⩽440) groups. Since TOEFL scores become

Table 3. Large score gains and large score attrition

	NERS July⇨Feb. TOEFL (459⇨457.5) Standard deviation = 24.5 Mean score increase = -1.5	ERS July⇨Feb. TOEFL (438⇨454) Standard deviation = 21 Mean score increase +16
Score gain ≥+1SD (all students)	13% Male = 11% / Female = 14.5%	18% Male = 14% / Female = 22%
Score attrition ≥-1SD (all students)	16% Male = 16.5% / Female = 15.5%	14% Male = 11% / Female = 17%

Table 4. Large score gains and large score attrition (high and low groups)

	Low group (TOEFL <440)		High group (TOEFL ≥441)	
	NERS (n=47)	ERS (n=34)	NERS (n=172)	ERS (n=42)
TOEFL change (July⇨Feb) points/SD	415⇨421 = +6 SD=27	420⇨445= +25 SD=22	471⇨467.5=-3.5 SD=23.5	453⇨462=+9 SD=19
TOEFL change (male and female)	M=F (+6)	Male =+20 Female=+30	Male=-6 Female=-1	Male=+8 Female=+10
Score gain ≥+1SD (all students)	9%*	27% Male=24% Female=29%	13% (M=F)	19% Male=11% Female=25%
Score attrition ≥-1SD (all students)	28% Male=21% Female=35%	3%*	15% Male=16% Female=14%	9.5%*

Note: *Percentages computed from 3 or fewer data points

increasingly difficult to improve as scores increase, as we might expect, overall score gains were better in the low group (LG) than the high group (HG). However how do ERS and NERS within these groups perform? Results indicate that in both the high and low groups, ERS out perform NERS significantly; LG ERS averaged +25 points against LG NERS +6 – a difference of 19 points gained. Similarly HG ERS gained +12.5 points over their NERS counterparts. For score changes both, within groups and across groups females out performed males; only amongst low group non-readers did men and women perform equally.

Looking at score gains ≥+1SD, ERS in both low and high groups showed a strong tendency towards large score gains (27% and 19% respectively). By gender, more LG males and females achieved large score gains (24% and 29% respectively) but in the HG such gains were only apparent in females (25%); high group males made far fewer large score gains than we would expect (11%).

At the ≥-1SD level, LG NERS showed a strong tendency towards score attrition (28%) while HG NERS were normally distributed (15%) suggesting a lack of reading may not be a factor in score attrition in the high group but is a *significant* factor in the low group. By gender LG NERS females large score attrition was twice as much (35%) as we might expect from a normally distributed sample, and considerably higher than males (21%). This suggests that females may have more to lose by not reading, particularly those with lower initial TOEFL scores. Too few ERS experienced score attrition ≥-1SD making meaningful gender analysis impractical.

Annual comparison

ERS and NERS score increases and pages per point

Score increases for the year (April 2006-February 2007) are described in Figure 1. This shows that despite poor second semester scores (in many cases score attrition), on a year long basis, irrespective of ER, students' average TOEFL scores *did* increase. The highest annual score gains were made by students who read in both semesters (ERS/ERS = +33.5 points) and the lowest by those who did no reading (NERS/NERS = +13.5 points) – see Figure 1 note for an explanation of this notation. The ERS/ERS students thus gained 20 TOEFL points over their NERS/NERS counterparts. On average ERS/ERS read 365 pages during the year meaning approximately 18 pages of reading were required to raise one TOEFL point.

Even students reading in one semester made TOEFL gains above non-readers (+17.5 and +22 TOEFL points) over the year with larger gains apparent in the semester in which reading was carried out. These two points' gains translate respectively to 49 and 19 pages/TOEFL point raised above the NERS students (see Table 5). The score losses (-1.5 TOEFL points) in the second semester in NERS/NERS and ERS/NERS may have been caused by examination fatigue but only serve to highlight the motivational benefits of extensive reading.

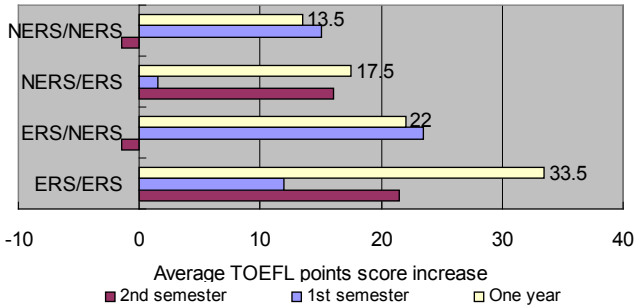

Figure 1. Mean annual TOEFL score increases (first and second semesters)

Note: The notation ERS/ERS used here and in Figures 1 and 2 refers to students who read in the first semester and second semester. Similarly NERS/ERS refers to students who did not read in the first semester but did read in the second.

High and low groups

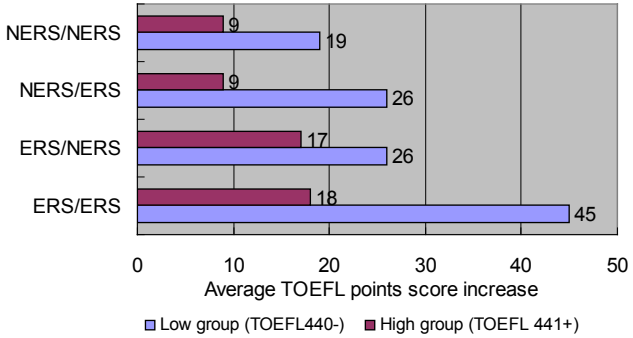

Figure 2. Mean TOEFL score increases (high and low groups)

Figure 2 describes the year long score increases for the high and low groups. As in Figure 1, over the year, students' TOEFL scores did improve. However in both groups the best gains were made by students reading in both semesters (low group ERS/ERS = +45 points and high group ERS/ERS =+18 points). These were +26 and +9 TOEFL points above non-reading counterparts and correspond respectively to pages/TOEFL point raised of 14 and 41 pages. Students reading in one semester made more modest gains. In the low group, single semester readers gained 7 points above non-readers approximating to 24-28 pages/TOEFL point raised, while in the high group, readers gained between 0 and 8 points above non-readers indicating that 20 pages of ER are required to raise one TOEFL point. The less beneficial effect of ER in the high group is emphasized by the single TOEFL point high group ERS/ERS made over ERS/NERS students. These figures for pages read per TOEFL point raised are summarized in Table 5 below.

Discussion

The results from this study suggest that extensive reading can help to raise TOEFL scores amongst poorly motivated students in an L1 environment. On both a semester and annual basis, these benefits seem to be not only in terms of score gains but also in helping to reduce score attrition. Students who read extensively are more likely to make

Table 5. Pages per one TOEFL point raised

	Studentcategory	Pages read	TOEFL points gained in one year	TOEFL points raised above control (NERS/NERS)	Pages per TOEFL point
All Students	NERS/NERS	0	13.5	0	-
	NERS/ERS	196	17.5	4	49
	ERS/NERS	164	22	8.5	19
	ERS/ERS	364	33.5	20	18
Low group	NERS/NERS	0	19	0	-
	NERS/ERS	196	26	7	28
	ERS/NERS	168	26	7	24
	ERS/ERS	360	45	26	14
High group	NERS/NERS	0	9	0	-
	NERS/ERS	168	9	0	-
	ERS/NERS	156	17	7	20
	ERS/ERS	372	18	8	41

large score gains (particularly females) than those who do not read and also experience fewer large score losses (though anomalously females may have more to *lose* by not reading than males). Initial TOEFL proficiency may also be a predictor in evaluating the relationship between ER and TOEFL score increases. The results here show that students starting with *lower* proficiency (i.e. under TOEFL 440) benefit more from ER than those starting above this score. Females however appear to benefit from ER whatever their initial TOEFL proficiency, while males only benefit at the lower initial proficiency level.

Explaining the reasons for such male and female differences may relate to the different ways reading is processed. According to Green and Oxford (1995), males and females use different approaches to language learning and this may have had an effect on the selection, reading and processing of reading materials. These different approaches relate to a wide range of strategies including skim reading, language review, summaries, asking for help from teachers and time scheduling which are all used more often by females than males. Since ER introduces a fast reading format, adopts a global reading approach (i.e., its goal is to review and summarizes what is already known rather than introduce new language), and requires good individual time management for book selection (and return), it suggests that female students may be better suited to ER than male students are. Although not investigated directly, time management and asking for help may have been important to the presumed ER gains in TOEFL scores as reading deadlines were set and book lending/return times were limited. Moreover, in most cases enquiries about ER in the PE programme were made by females rather than males. Other research supporting the notion that females are better suited to ER indicates that female students also tend to attach a higher pre-reading value to extensive reading and have a greater belief in its ability to assist their TOEFL scores than males (Cramer, Ascough, Williams, & Loucky, 2007, p. 987). Such value is likely to increase motivation and reapplication of learning encountered in ER during the TOEFL test itself. In the current research females were more inquisitive about the books' contents while males tended to choose whatever titles were available. Often this was fiction that was more accessible to females and this may have lowered males' motivation to read and reduced the opportunity to apply ER learnt skills and knowledge to the TOEFL context.

In explaining the apparent greater benefit ER gives students with lower initial TOEFL levels three possible mechanisms may be at work. Firstly, TOEFL score improvement is not linear but geometric with greater effort being required to raise one TOEFL point amongst students with higher initial scores than those with lower initial scores. It is quite possible that reading 4.8 books per semester does not represent sufficient effort at the higher TOEFL level than at the lower one. Secondly, since many low group students had initial TOEFL scores under 400, it is possible they were actually *learning* while reading whereas the higher level students were only reviewing and consolidating language – as we would expect from ER. In this way students with lower initial scores would have derived wider benefit from ER and this may have been reflected more in their average score gains. A third possible explanatory factor relates to

the dominance of fiction over non-fiction titles held in the reading salon itself. By having few non-fiction titles (a good means to develop topic specific language relevant to the TOEFL test), higher level students who would have already possessed the more general language offered in fiction titles may have been less able to develop their TOEFL scores through ER.

These observations show that a considered approach to ER orientation and usage, library management, staffing and book purchasing are very important to ensure the most suitable resources are available. As well as the types of studies students are taking, the gender make up of the student body and their overall language capability and motivation need to be considered. Moreover, ways to overcome the inertia students feel towards reading need to be found. Rather than being motivational the TOEFL 500 requirement at YCU made many students feel disillusioned, particularly less able ones. In this sense ER gave students something they could handle and motivated them. This is clearly illustrated in the lower drop out rates from PE in students who read (11%) against those who didn't (20%). Mixed ability groups as opposed to level based stratification may, through peer exchange, also help in informing the higher ability students of the possible TOEFL score gains ER may promote.

While the semester results help to inform us of how to manage ER resources, the year long data help to form a clearer idea of how much ER might be able to assist TOEFL scores. As Table 5 shows students can expect to need to read between 14 and 49 pages in order to raise their TOEFL score by one point. Although these figures correlate to those found in previous studies (Mason, 2006), such a wide range may indicate that other factors unrelated to ER are driving up TOEFL scores, and that the gains observed in this research are not necessarily attributable to ER. However, since 7 of the 9 pages per TOEFL point values shown in Table 5 are between 14 and 28 pages/TOEFL point raised, it seems sensible to suggest this range as being representative of the effect ER might have on TOEFL scores. If we could arbitrarily extend the mid-point of this range (i.e. 21 pages/TOEFL point) to a 70-hour English programme (about the same as PE) and devote half its instruction time to ER as Lee (2007) suggests, we might, given the 70 minutes a typical student needs to read a level appropriate extensive reading book (Cramer et al., 2007), expect score gains of as much as 57 points per *semester*. Informing students of such data may be a very powerful way to overcome the inertia described above.

Finally some caution should be read into the findings described here. Firstly, for convenience it was assumed that all students (i.e. both readers and non-readers) had similar study habits in and out of the classroom. Given the "virtuous circles of learning" (Nuttall, 1982, p. 168) that ER promotes this assumption may be unreasonable since students who read would have been more likely to undertake extracurricular and other language practice in tandem with ER. It may thus be these other activities that contribute to some part of the TOEFL score gains in student readers and not the ER alone. Secondly, the compulsory nature of reading in the current study offers no guarantee that students actually read the books they borrowed. Nevertheless similar assumptions have been made in previous smaller scale studies linking ER and TOEFL scores and since the current

study was both larger in scale and more longitudinal in attitude such weaknesses will have been minimized in the results and analysis described here.

One of the biggest problems in measuring the effect of reading extensively on TOEFL score changes seems to be reading itself. Even amongst willing volunteers research has indicated that reading freely is something students do not take to easily (Constantino et al., 1997; Mason, 2006), and there is a general belief that ER may not be relevant to TOEFL proficiency (Holden, 2003). The current study also came up against such inertia. We should thus consider how we present reading to students and how they view ER and TOEFL in a programme like PE. Teachers will need to invest time to encourage students, and students will need to be aware of the effort required to make ER an enjoyable and beneficial activity.

Conclusion

The significant changes in TOEFL scores associated with ER, described in this research, along with the importance of these scores to tertiary education suggest that ER can play a role beyond simply giving students improved scores on standardized language tests. With higher scores, not only can students can gain access to more educational and employment opportunities but, universities can secure new ways to promote their academic functions and develop improved language learning programmes.

David Williams is a lecturer at Yokohama City University and has lived in Japan for more than 15 years. His current research interests include the benefits extensive reading has on standardized tests scores, and student motivations behind study abroad. <davidokj@yahoo.com>

References

Constantino, R., Lee, S. Y., Cho, K. S., & Krashen, S. (1997). Free voluntary reading as a predictor of TOEFL scores. *Applied Language Learning 8* 111-118.

Cramer, T., Ascough, R. K., Williams, D., & Loucky, J. P. (2007). Extensive reading forum: Motivation and community through extensive reading. In K. Bradford-Watts (Ed.), *JALT2006 Conference Proceedings*. Tokyo: JALT.

Day, R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language 7* 541-55.

Elley, W. B. (1991). Acquiring literacy in a foreign language: The effect of book based programs. *Language Learning 41* 375-411.

Gradman, H., & Hanania, E. (1991). Language learning background factors and ESL proficiency. *Modern Language Journal 75*(1) 39-51.

Green, J. H., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly 29*(2) 261-295.

Holden, W. R. (2003). Student attitudes towards graded reading: A preliminary investigation. *Bulletin of Hokuriku University 27* 145-158.

Krashen, S. (2004). *The power of reading: Insights from research*. Portsmouth: Heinemann.

Lee, S. Y. (2007). Revelations from three consecutive studies on extensive reading. *RELC Journal 38*(2) 150-70.

Mason, B. (2006). Free voluntary reading and autonomy in second language acquisition: Improving TOEFL scores from reading alone. *International Journal of Foreign Language Teaching 2*(1) 2-5.

Nation, P. (1997). *The language learning benefits of extensive reading*. [Online] Available:<http://www.jalt-publications.org/tlt/files/97/may/benefits.html>

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

Takase, A. (2008). SSR for reluctant learners to read. *Kinki University English Bulletin 2* 19-36.

Tsang, W. K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics 17* 210-233.

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) 130-157.