Determining the Crucial Characteristics of Extensive Reading Programs: The Impact of Extensive Reading on EFL Writing

by Sy-ying Lee, Ph.D. and Ying-ying Hsu

Sy-ying Lee, Ph. D., Professor of the Department of Foreign Language and Applied Linguistics, Taiwan. Research interests and publications cover areas of writing apprehension and second language writing, extensive reading, storytelling and sustained silent reading as paths to a more successful second language acquisition.

Ying-ying Hsu, MATESOL, Senior Lecturer of English at Nanya Institute of Technology. She collaborated with Syying Lee in a three year longitudinal project tracing how in-class sustained silent reading effects English acquisition and has had several publications coauthored with Prof. Lee.

Abstract

This one-year study examined the impact of in-class extensive reading or sustained silent reading on writing with a group of Taiwanese vocational college students. These students had been less successful in academics, including English. While many researchers and practitioners believe that less proficient ESL/EFL students need more direct instruction, sustained silent reading has been gaining support from research. The design attempted to avoid the weaknesses in the design of previous studies by having a longer duration, an appropriate comparison group, providing more access to books, and requiring less accountability. Subjects devoted part of the class time to in-class reading and followed the same writing curriculum as the comparison group did. Pre and post essays were graded following Jacobs et al.'s (1981) measurement of writing, which included five subscales: content, organization, vocabulary, language use, and mechanics. Results showed significant differences in gains on all subscales in favor of the experimental group.

Some scholars have hypothesized that writing style, the special language of writing, is acquired, or subconsciously absorbed, through reading (Krashen, 1984, 2004; Smith, 1988, 2004). The language of writing, it is argued, is too complex to be consciously learned, and there are compelling case histories of those who developed high levels of competence in the written language through reading alone, without instruction (see review in Krashen, 2004).

Correlational studies confirm that more reading is related to better writing in both first and second language research (Alexander, 1986; Applebee, 1978; Applebee, Langer, & Mullis, 1986; Lee, 1995, 2001, 2005; Lee & Krashen, 1996; Lee & Krashen, 2002; Huang, 1996; Gradman & Hanania, 1991; Janopoulos, 1986; Kaplan & Palhinda, 1981; Mason, 2004).

A few classroom experiments have been conducted to determine whether self-selected, extensive reading in the classroom effectively enhances writing ability among EFL and ESL learners (Elley & Mangubhai, 1983; Elley, 1991; Tsang, 1996; Lai, 1993; Hafiz

Research Index • Teacher-to-Teacher Index • Submission Info • Contact Us • Subscription Info

& Tudor, 1990; Tudor & Hafiz, 1989; Mason & Krashen, 1997). Four of these studies (Tsang, 1996; Lai, 1993; Hafiz & Tudor, 1990; Tudor & Hafiz, 1989) examined the effect of a reading program on the writing performance of subjects who learned English as a foreign or second language and who had a similar educational background, and in all four of these studies, students were evaluated on descriptive writing, writing on topics such as "My family," "Coming to School," etc.

These studies have, in general, produced positive results, with readers outperforming comparison students on measures of language use: percentage of syntactic and semantically acceptable sentences (Hafiiz and Tudor, 1989; Tudor and Hafiz, 1990; Tsang, 1996); spelling and fluency (Hafiz and Tudor, 1989; Tudor and Hafiz, 1990; Lai, 1993); accuracy, percent of error-free T-units (Lai, 1993); variety of vocabulary used (Tudor and Hafiz, 1990); content (Tsang, 1996); and overall impression (Lai, 1993; Tsang, 1996). But readers did not significantly outperform comparisons on spelling in one study (Tsang, 1996), on vocabulary in two studies (Tudor and Hafiz, 1990; Tsang, 1996), and on organization in another (Tsang, 1996).

It is surprising that the results were so positive, as all the studies suffered from serious design flaws due to practical constraints:

Participants in previous studies had little to read. Tudor and Hafiz (1989) provided only 104 books for 16 students, Hafiz and Tudor (1990) provided 106 books for 25 students, in Lai (1993), each class of 20 students had 40 books, and Tsang's students (Tsang, 1996) were required to read only eight books.

In each study, students were required to make thorough reports on what they read, which may have extinguished some of the pleasure of reading. Students were required to make oral reports (Hafiz and Tudor, 1989; Tudor and Hafiz, 1990; Lai, 1993) or had to fill out "review forms" which were graded for "details and persuasiveness" (Tsang, 1996).

The duration of all previous studies was short, ranging from four weeks (Lai, 1993) to three months (Hafiz and Tudor, 1990).

Our hypothesis is that access to more reading material, less accountability, and a longer duration would show a larger and more consistent impact of self-selected reading on measures of writing.

Procedure

The subjects were eighty-six (86) students in the Department of Applied Foreign Languages at a vocational college in Taiwan. Mandarin was the language of instruction in nearly all courses. All students were in the third year of a five-year program in vocational education, were 17 and 18 years old, and had studied English previously for five years. These students were in general less successful than other university students in Taiwan, with most having failed the nation-wide senior high entrance examination. There were forty-three (43) students in the experimental, extensive-reading, group and forty-three (43) in the comparison group. The number of subjects, however, varied slightly in each analysis because some students did not complete all the tests.

All subjects were taking a general English class and a writing class. All subjects had been taking general English classes since their first year of their five-year college education and started the writing class in their third year, the year this project was conducted. The study was conducted as part of the general English class, which met twice a week, one session for 100 minutes and one for 50 minutes. The duration of the study was one academic year, two semesters, much longer than the duration of any of the studies discussed above.

The General English Class. In the general English class, the comparison group had traditional instruction, which included vocabulary and grammar study, and reading passages followed by comprehension questions. Some time was also devoted to group discussion. Quizzes, a mid-term, and final examinations were given as required by the

curriculum. The experimental group also followed the same traditional instruction for 100 minutes per week, but did self-selected reading for the 50-minute period each week. The experimental students were required to take the same quizzes and examinations as the comparison group. Both groups had the same instructor, the second author.

The Writing Class. Both groups met twice a week, one session for 100 minutes, the other for 50 minutes and both used the same textbook, *First Steps in Academic Writing* by Ann Hogue. Different instructors taught the two writing classes, but both instructors followed the chapters in the text, aimed at acquainting students with the composing process, providing practice producing basic sentence structures, and helping students develop grammatical and mechanical skills. Both groups were required to write three to four essays each semester. No changes were made in the usual practice in the writing class for the purposes of this study.

Reading Material. The instructor provided about 530 graded readers and titles from the Penguin and Oxford reading series, more than 12 books for each student, a ratio considerably higher than in previous studies. Penguin Readers are divided into six levels ranging from 300 headwords to 3000 headwords. Oxford readers also have six levels containing headwords up to 3000. Experimental group students were allowed to choose materials to read according to their own interests and language proficiency level, and it was suggested that they read at least one book per week.

Experimental students were asked to fill out a reading log recording how many pages they read and how much time they spent reading. In addition, students were encouraged to write a brief reflection paragraph or summary after they finished reading each book, either in English or Chinese, far less accountability than was required of participants in the studies reviewed above. Those who wrote reflections and summaries typically wrote only a few sentences. The instructor made suggestions on book selection based on students' comments and requests in the logs.

Measures. Two writing samples, written without feedback and revision, served as pre- and post-tests. The pre-essays were the first assignment at the beginning of the first semester and the post-essays were collected at the end of the academic year. In both cases, students were asked to do descriptive writing, with "The Moon Festival" and "Your Summer Vacation" as the topics for the pre- and post-tests. Two raters read the writing samples, both senior professors of English with many years of experience in grading compositions for the nation-wide university entrance examination. Essays were evaluated based on criteria established by Jacobs, Zinkgraf, Wormuth, Hartfield, and Hughey (1981): content (points awarded ranged from 13-30), organization (7-20), vocabulary (7-20), mechanics (2-5), and language use (5-25).

Following Jacobs et al., overall impression was calculated by totaling the components, but with different weightings for the subscales, with more value placed on content (13-30) and structure (7-20), and less on mechanics (2-5). The total number of words written was used as a measure of fluency. Inter-rater reliability based on the pre-essay scores was .87 using Kendall's W.

In addition, grades for the four examinations for the general English class were collected to ensure that both groups were at a similar English proficiency level before treatment (CONTENT EXAM 1, based on material covered the previous semester), and also to observe how extensive reading affected academic performance in the class (MIDTERM, CONTENT EXAM 2, given at the end of the first semester, and CONTENT EXAM 3, given at the end of the second semester).

Moreover, an attempt was made to evaluate subjects' reactions to the reading program by questionnaire and by noting the number of pages they read each semester. We assume that an increase in the amount students read reflects a positive attitude.

Results

The two groups had very similar grades on their final examination for the general English class in the previous semester (CONTENT EXAM 1) before treatment (experimental = 81.4, SD = 9.9; control = 80.5, SD =.7; t = .37; p = .71). There were also no significant differences on pre-test measures of vocabulary and reading[1].

For the comparisons of the measures related to writing, a multivariate method, Hotelling's T- Square, was used. This method allows a comparison of the mean values of two groups with more than one pair of dependent variables. In this study, six pairs of variables were compared: fluency (number of words written), content, organization, vocabulary, mechanics, and language use. Results indicated that the groups were not significantly different on the pre-essay (Hotelling's T = .12, p = .16), and the individual p values obtained from the multivariate analysis showed that there were no differences between the two groups on all subcomponents, including fluency (Table 1).

Table 1. Scores of the Pre Essays and Multivariate Analysis between Groups

	Exp	Con	Mean Diff.	p
Fluency	98.47 (34.86)[a]	97.76(30.14)	.71	.92
Content	16.65 (2.39)	16.65 (2.36)	.00	.99
Organization	9.79 (1.52)	10.04 (1.52)	.25	.46
Vocabulary	9.76 (1.52)	10.23 (1.46)	.47	.15
Language Use	9.76 (2.63)	10.65 (2.28)	.90	.10
Mechanics	2.99 (.54)	3.12 (.69)	.13	.33

N = 43 for the experimental group and 42 for the control group
a. The numbers in the parentheses denotes the standard deviations for raw scores

At the end of the second semester, the experimental group significantly outperformed the control group on the post-essay (Hotelling's T = .62, P < .000), performing better on all subscales (Table 2) and made larger gains than the control did (Table 3). The inter-rater reliability of the post-essay using Kendall's W was .82. It is reasonable that reliability was lower in the post-essay than on the pre-essay, because of the greater variability in scores caused by the greater gains made by the experimental group.

Table 2. Scores for the Post Essay and Multivariate Analysis between Groups

	Exp	Con	Mean Diff.	p
Fluency	123.02(37.06)	82.17(34.86)	40.85	.000**
Content	19.34 (2.46)	16.88 (2.73)	2.46	.000**
Organization	12.79 (1.87)	10.54 (2.09)	.2.26	.000**
Vocabulary	11.62 (1.44)	10.10 (1.74)	1.52	.001**
Language Use	12.73 (2.51)	10.78 (2.52)	1.95	.001**
Mechanics	3.33 (.61)	3.02 (.61)	.31	.026*

Note: ES = effect size.
N = 41 for both groups in the post essay test.

Table 2 also presents effect sizes, based on mean scores, for all sub-measures of writing used in this

1 In this one year project, subjects were also tested on vocabulary using Nation's measure (1990) and reading using Mason's cloze test (2003). Details were described in detail in Lee (2007).

study. The effect size shows the magnitude or impact of a treatment. According to Cohen (1988) an effect size of 0.2 represents a small effect, 0.5 a medium one, and 0.8 a large effect. The results in this study show effect sizes mostly larger than 0.8 in favor of the experimental group, meaning that the experimental group was clearly superior to the control group on nearly all of the subscales in Jacob et al's writing measures.

Table 3. Gains on Each Subscale for Both Groups

	Exp	t / p value	Con	t / p value	ES
Fluency	23.05(36.04)	4.10/.000**	-14.4 (31.84)	-2.92/.006**	1.11
Content	2.62(2.75)	6.10/.000**	.28 (2.81)	-.62 /.54	.85
Organization	2.94(1.90)	9.86/.000**	.50 (1.89)	1.67/ .10	1.30
Vocabulary	1.78(1.55)	7.34/.000**	-.11(1.77)	-.40 / .69	1.15
Language Use	2.85(2.58)	7.07/.000**	.18(2.47)	.45 / .66	1.07
Mechanics	.33(.57)	3.73/.001*	-.11(.67)	-1.06 /.30	.72

Note: N = 41 for the experimental group and 40 for the control group

Table 4. Comparisons of the Course Required Examinations for Both Groups

		CONTENT-1 (before treatment)	MIDTERM (during the first semester)	CONTENT-2 (after one semester)	CONTENT-3* (after one year)
Con	Mean (SD)	81.35 (9.87)	71.09 (11.30)	70.43 (11.43)	69.27 (11.27)
	N	43	43	43	42
Exp	Mean (SD)	80.51 (10.72)	75.63 (13.21)	72.13 (14.35)	74.74 (11.89)
	N	43	43	43	43

The two aspects for which the effect sizes were below .8 (language use and mechanics) still represent a modest victory for extensive reading over traditional instruction on writing.

The reading group improved significantly on all aspects, while the control group did not, a rather frustrating outcome for those who believe in the efficacy of traditional instruction. Table 3 presents the results of t-tests comparing pre and post-essays for each group on each subscale, as well as effect sizes based on gain scores.

Note that the control group even declined significantly on the indicator of fluency (number of words written), and regressed slightly on the measures of content, vocabulary and mechanics. What is even more intriguing is that the reading groups outperformed the control group on the course required examinations (Table 4).

Table 4 shows that the experimental group did better on all content examinations after the pre-tests, and the difference was statistically significant on the final test given at the end of the year ($t = 2.22$, $p = .03$; $d = .84$)[2].

In addition, the experimental students reported reading more pages in the second semester than they did in the first semester (525.36 pages vs. 357.52 pages, $t = 5.93$, $df = 41$, $p < .000$), suggesting that they enjoyed doing the reading.

2 The lower grades on the examinations following CONTENT TEST 1 was most likely due to the fact that 30% of the items on these tests were on material not covered in the textbook. These items were drawn from the GEPT practice test. The GEPT (General English Proficiency Test) was developed by the Language Training and Testing Center in Taiwan, an organization responsible for test development and awarding certificates indicating one's English proficiency level. The finding that the reading group performed better in these examinations suggests that extensive reading works no less effectively, or even more effectively, than instruction in preparing students to take GEPT.

Students in the reading section also filled out a brief questionnaire, in Mandarin, at the end of the first semester for us to evaluate the program and seek ways to improve the program in the second semester. They were asked if they felt that they had improved in different aspects of language competence, and were also asked what activities they felt needed to be strengthened or included in the class.

Table 5. Students' Reflections on In-Class Extensive Reading after One Semester

Aspects felt improved	N	%
a. reading ability	31	72.1
b. reading rate	30	69.8
c. vocabulary	17	39.5
d. grammar	7	16.3
e. writing	7	16.3
Activities felt needed		
a. more grammar teaching	16	37.2
b. more story analysis	20	46.5
c. more books	33	76.7
d. more discussions	9	20.9
e. others	3	7

As presented in Table 5, students clearly felt they had improved in reading and vocabulary, but were less certain about writing and grammar. Our results, however, confirm that they did indeed improve in writing after one year of reading, as previously shown. In addition, most students recommended more books, which confirms that they were enthusiastic about reading.

Some, however, felt a need for more traditional instruction, grammar and story analysis.

At the end of the program, an open-ended questionnaire was distributed to determine if students still felt that the in-class reading program helped. There were some unexpected results (Table 6). Students reported that they felt they struggled less in figuring out the meanings of unfamiliar words, i.e. they developed better strategies for guessing word meanings without consulting the dictionary, taking advantage of context as an aid to comprehension. A few students mentioned that reading helped them with their listening comprehension, that they began to think in English when reading, and that they even progressed on pronunciation.

Table 6. Students' Reflections on Extensive Reading after One Year

Aspects felt improved	**N**	**%**
a. vocabulary	16	37.2
b. reading comprehension & speed	15	34.8
c. strategies for guessing unknown words and expressions	12	27.9
d. grammar	9	20.9
e. thinking in English	1	2.3
f. listening in English	1	2.3
g. pronunciation	1	2.3

Only the reading group did the questionnaire. The results are therefore only suggestive.

Table 7. Comparison to Tsang (1996) Using Effect Sizes

	this study	**Tsang**
organization	1.32	0.32
content	0.96	0.75
vocabulary	1.15	0.32
language use	1.15	0.63
spelling/mechanics	0.75	0.14

Discussion

The goal of this study was to determine whether an improved design would result in a greater impact of self-selected reading on writing. This study, therefore, provided far more access to books, lasted longer, and did not require extensive and potentially stressful post-writing activities.

In this study, the practice of extensive reading was the only activity in which experimental subjects

differed from comparison subjects. All did the same number of writing assignments in writing class, all followed the same regular syllabus, and there were no pre-test differences. As the results indicated, reading clearly had a strong effect in this study, with readers showing a clear and strong superiority in fluency, as well as in all the subscales of the Jacobs et al. writing assessment tool.

Table 8. Comparison to 4 Studies Using Statistical Significance

	this study	Tsang	Hafiz & Tudor	Tudor & Hafiz	Lai
organization	sig	ns			
content	sig	sig			sig
vocabulary	sig	ns	ns	ns	
language use	sig	sig	sig	sig	
spelling/mechanics	sig	ns		sig	sig
fluency	sig		sig	sig	sig

Sig = readers significantly better than comparisons

Ns = readers not significantly better than comparisons

We were able to compare our results directly with those of Tsang (1996), because the same measure was used in both studies. Students in our study made greater gains, as shown by a comparison of effect sizes (Table 7).

It was not possible to compare our results using effect sizes with the other three studies, as measures were often only roughly comparable across studies. Students in our study, however, did as well as or better than students in all four studies, as shown in Table 8, which uses a cruder measure, statistical significance rather than effect size.

Table 9. Total Duration, Total Time in Reading

Study	Duration	Total Time
The present study	30 weeks	25 hours
Tudor & Hafiz	3 months	42 hours
Hafiz & Tudor	23 weeks	90 hours
Lai	4 weeks	50 hours
Tsang	24 weeks	

These results are consistent with our hypothesis that longer duration, increased access and/or less accountability make a difference. They are also consistent with Lee (2007) who found that these were the key factors of a more successful extensive reading program after three consecutive studies with Taiwanese university students. Our results, of course, do not allow us to specify which of these aspects is crucial or the most important.

This study confirms that longer term studies produce better results, consistent with Krashen (2004). It is of interest, however, that the duration of our study was longer, but the total time devoted to reading was not. In fact, our subjects devoted less total time to reading (25 hours; 50 minutes per day for 30 weeks) than subjects did in any of the studies reviewed here (Table 9). This could mean that distributed rather than massed reading sessions are more efficient, that is, shorter sessions more spread out over time.

The results are also consistent with the studies that show that greater access to reading material results in more reading, and in turn better achievement on tests of reading and writing development (Krashen, 2004). Our students had access to more than 500 titles, about five times as many as Tudor and Hafiz' subjects were provided with and more than ten times as many as Lai's. In terms of books per student, our ratio was 12.4 to 1 (534 titles/43 students), far greater than the ratio in any of the previous studies.

Our results extend the research on self-selected reading to vocational college students, providing evidence for the reliability of the efficacy of extensive reading. One possible flaw in our design was the fact that two different

topics were used as prompts in the pre- and post-essays. It is possible that students react differently to different prompts. Few studies, however, have used the same topics for both pre- and post-tests because of the concern about student familiarity with the topic, and both the pre-essay prompt,The Moon Festival, and the post-essay, Your Summer Vacation, were similar in that students were required to describe how they spent their time during the holiday and vacation. Neither asked for special background knowledge and terminology to complete the task. Most important, both the experimental and comparison group wrote on the same topic.

Our results do not, of course, demonstrate that reading is the only source of competence in writing, but it is doubtful that formal instruction makes a substantial contribution, evidenced by the results of the comparison group as well as the fact that the written language is so complex. The results are, however, consistent with the hypothesis that the source of good writing style, the vocabulary, syntax and discourse structure of the written language, is reading.

Acknowledgement

The support of the National Science Council, The Republic of China, under grant NSC 94-2411-H-305-009, is gratefully acknowledged.

References

Alexander, F. (1986). *California assessment program: Annual report.* Sacramento: California State Department of Education.

Applebee, A. (1978). Teaching high-achievement students: A survey of the winners of the 1977 NCTE Achievement Awards in writing. *Research in the Teaching of English, 1*, 41-53.

Applebee, A., Langer, J. & Mullis, I. (1986). *The writing report card.* Princeton, N.J.: Educational Testing Service.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ.: Lawrence Earlbaum Associates.

Elbow, P. (1973). Writing without Teachers. New York: Oxford University Press.

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

Elley, W. & Mangubhai, E. (1983). The impact of reading on second language learning. *Reading Research Quarterly, 19*, 53-67.

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

Hafiz, F. & Tudor, I. (1990). Graded readers as an input medium in L2 learning. System, 18(1), 31-42.

Huang, I. C. (1996). The relationship of free reading and writing proficiency in second language acquisition. Proceedings of the 13th ROC TEFL, Oct, 1996, pp. 133-143.

Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfield, V. F. & Hughey, J. B. (1981). *Testing ESL composition: A practical approach.* Rowley, Mass: Newbury House.

Janopoulos, M. (1986). The relationship of pleasure reading and second language writing proficiency. TESOL Quarterly, 20, 763-768.

Kaplan, J. & Palhinda, E. (1981). Non-native speakers of Englilsh and their composition abilities: A reveiew and analysis. In W. Frawley (Ed.), Linguistics and literacy (pp. 425-457). New Yort: Plenum Press.

Krashen, S. (2004). The Power of Reading (2nd Ed.). Englewood, CO: Libraries Unlimited.

Lai, F. K. (1993). The effect of a summer reading course on reading and writing skills. System, 21(1), 87-100.

Lee, S. Y. (1995). *A cross-cultural validation of The Reading Hypothesis: The relationship of pleasure reading to writing proficiency and academic achievement among Taiwanese senior high school students.* Doctoral dissertation, University of Southern California, Los Angeles, CA., U.S.A.

Lee, S. Y. (2001). *What makes it difficult to write: Theory, research, and implications.* Taipei: Crane publishing Co., Ltd.

Lee, S. Y. (2005). Facilitating and inhibiting factors on EFL writing: A model testing with SEM. *Language Learning*, *55*(2), p. 335-374.

Lee, S. Y. (2007). Revelations from Three Consecutive Studies on Extensive Reading. *Regional Language Center (RELC) Journal, 38*(2), 150-170.

Lee, S. Y. & Krashen, S. (1996). Free voluntary reading and writing competence in Taiwanese high school students. *Perceptual & Motor Skills*, *83*, 687-690.

Lee, S. Y. & Krashen, S. (2002). Predictors of success in writing in English as a Foreign Language: Reading, revision behavior, and writing. *The College Student Journal, 36*(4), p. 532-543. USA.

Mason, B. (2003). *Sufficiency of extensive reading on the development of grammatical accuracy.* Doctoral dissertation, Temple University, Tokyo.

Mason, B. (2004). The effect of adding supplementary writing to an extensive reading program. *International Journal of Foreign Language Teaching*, *1*(1), 2-16. Available at http://www.tprstories.com/ijflt/IJFLTFall05.pdf

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

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

Tsang, W. K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics*, 17(2), 210-233.