

Beniko Mason* and Stephen Krashen**

**International Buddhist University, Habikino-shi, Osaka, Japan*

***University of Southern California, Los Angeles, CA, U.S.A.*

Three experiments confirm the value of extensive reading in English as a foreign language (ELF). In extensive reading, students do self-selected reading with only minimal accountability, writing brief summaries or comments on what they have read. In Experiment 1, "reluctant" EFL students at the university level in Japan did extensive reading for one semester. They began the semester far behind traditionally taught comparison students on a cloze test, but nearly caught up to them by the end of the semester. In Experiment 2, extensive readers outperformed traditionally taught students at both a prestigious university and a two-year college. In Experiment 3, extensive readers who wrote summaries in English made significantly better gains on a cloze test than a comparison class that devoted a great deal of time to cloze exercises. Gains made by extensive readers who wrote in Japanese were greater than comparisons, but the difference was not significant. Those who wrote in Japanese, however, made gains superior to both groups on a measure of writing and in reading speed. © 1997 Elsevier Science Ltd. All rights reserved.

Despite the growing amount of research supporting the use of extensive reading for improving second language competence (e.g. Elley and Mangubhai, 1983; Tudor and Hafiz, 1989; Hafiz and Tudor, 1989; Elley, 1991; Pilgreen and Krashen, 1993; Cho and Krashen, 1994, 1995a,b; Constantino, 1994), many teachers are still uncertain about how effective it is. This series of studies addresses several concerns about extensive reading.

1. Because some of the studies supporting extensive reading have been done in the second language environment, is it viable in a foreign language acquisition environment? Will foreign language students who do extensive reading do as well as those in traditional classes?
2. Some maintain that extensive reading will only benefit more successful and more motivated students, and will not help those who are unmotivated and who have not done well in language classes; these poor students, it is argued, lack the grammatical knowledge and vocabulary that is necessary for reading comprehension and enjoyment.
3. Can extensive reading be enhanced by the use of output/writing activities in the foreign language?

4. Is the effect of extensive reading reliable? As noted above, there is only a handful of studies supporting the use of extensive reading in second and foreign language acquisition.

Experiment 1: Extensive Reading And Reluctant Students of EFL

In this study, we investigate whether so-called "bad students" or failures in English as a foreign language (EFL) could improve with an extensive reading treatment.

Procedure

Subjects. Subjects were members of two intact EFL/reading classes at a women's university in Osaka, Japan. The comparison group consisted of second year students in the general education curriculum. The experimental class was designed for students who had failed EFL classes, termed a *Sai Rishu* (retakers) class. Students in this class were second, third and fourth year students. Both classes were held once a week for 90 min and were taught by the same teacher (B.M.).

Treatment. For the first semester of the year, both classes followed a traditional curriculum, which included reading selections, comprehension questions, vocabulary and grammar exercises. Students in the comparison class adapted well to this method, turning in assignments on time and doing well on exams. Attendance was nearly perfect. The *Sai Rishu* class was different: very few students turned in their homework, much of which was incomplete, and students were often late to class. Overall, attendance was poor, and one-third of the students dropped the class before the end of the semester. Test scores were very low. Because of this experience, a new approach was tried for the second semester with the *Sai Rishu* class.

While the comparison group continued with traditional instruction, the *Sai Rishu* class spent the second semester reading graded readers, both in class and as homework. About 100 books for students of EFL were purchased for the 30 students in the class. Students were required to read 50 books during the semester, and were also required to write short synopses and keep a diary in Japanese recording their feelings, opinions, and progress. Books ranged from the 600 to the 1600 word level (a 1600 word level book contains about 16,000 words). During class sessions, the teacher checked students' notebooks, discussed their reading with

them, and encouraged them. The 50 book requirement proved to be too ambitious, but by the end of the semester, some students had read over 40 books, and the average number of books read was about 30.

Measure. A 100-item cloze test, written at the sixth grade level, was given as a pre-test and post-test to both groups. The text was about a little girl's experience at the time of the atomic bomb in Hiroshima. Thus, subjects had some background knowledge related to the story. The test was constructed from a reading passage of about 1600 words, with every 10th word deleted, and all acceptable answers were considered correct, even if they were not identical with the original text. We allowed one hour for the test. Test-retest reliability, calculated from separate administrations of the test two weeks apart with a sample different from the subjects of this study, was 0.87.

Twenty subjects were randomly selected from each group for the study.

Results

Table 1 presents means for the pre- and post-test and gain scores. As expected, the comparison group outperformed the experimental group on the pre-test, but inspection of post-test scores shows that the experimental group nearly made up the gap between the groups.

The gains made by the experimental group were significantly greater than the gains made by the comparison group ($t = 2.269$, $df = 38$, $p < 0.025$).

Perhaps the most important and impressive finding of this study is the clear improvement in attitude shown by the experimental students. Many of the once reluctant students of EFL became eager readers. Several wrote in their diaries that they were amazed at their improvement. Their diaries also indicated that they understood the stories. Also of interest is our observation that students did not progress linearly from easy to harder books. Some students read easier books after reading some more difficult tests, and then returned later to harder books.

Table 1. Pre- and post-test scores

	Pre-test		Post-test		Gain	
	Mean	SD	Mean	SD	Mean	SD
Experimental	22.55	11.54	31.40	11.43	8.90	6.22
Comparison	29.70	8.23	33.05	8.24	4.35	6.47

Discussion

The clear gains made by the experimental group are quite consistent with previous reports of the positive effect of extensive reading on second language acquirers. Taken in isolation, however, these results of this study are, at best, suggestive. It is a study of only two classrooms, only one measure was used, and one of the experimenters taught both sections.

Experiment 2 was designed to establish the reliability of the effect of extensive reading with two additional groups, one from a prestigious university and one from a junior college. This time, the experimenter taught only the extensive reading section, and the treatment was lengthened to one academic year.

Experiment 2: Extensive Reading In A Prestigious University And A Junior College

Procedure

Subjects. Four intact classes of EFL students were used, a total sample size of 128 students, two classes from the English literature department at a highly regarded four-year university, and two from a junior college. The experimental classes in each institution read from graded readers; the students from the four-year college were able to move to authentic texts the second semester. Students read at their own level and at their own pace. Extensive reading was the main part of the course.

When students finished a book, they wrote a summary of the book in English (from a half to a whole page), then wrote an "appreciation" of the book in Japanese as well

as reflections on their reading comprehension. Students also chose one book each week and told a partner about the book in English. Thus, as in our previous study, there was some "accountability", but this accountability was low pressure.

Regular classes were taught the traditional way, focusing on direct teaching of reading comprehension and intensive reading of short, difficult, assigned passages, which necessitates a great deal of dictionary work.

Measures. All students were pre- and post- tested on the same 100-item cloze test used in Experiment 1. In addition, the experimental students wrote a summary of the first book they read, and repeated this procedure with a different book at the end of the academic year. These summaries were done without consulting the book and were rated by three native speakers. The judges had no knowledge of whether the papers they read were written at the beginning or end of the year, and were asked to rate each essay as "good", "average", and "not good", using any criteria they wished. In addition, experimental subjects also filled out a brief questionnaire during the final session.

Results

Cloze test. Table 2 presents pre- and post-test scores for all four groups.

As expected, the two-year college students had lower pre-test scores. It is clear from Table 2 that the experimental groups in both institutions made better gains than the regular students, and in both cases the differences were statistically significant (for university students: $t = 4.991$, $df = 72$, $p < 0.001$; for junior college students: $t = 5.035$, $df = 46$, $p < 0.001$).

Writing. Table 3 presents judgments for each of the three judges individually. In every case, more summaries were classified as "good" in the post-test than in the pre-test. To allow statistical analysis, "average" and "not good" categories were collapsed.

Table 2. Cloze test results

	n	Pre-test	SD	Post-test	SD	Gain
University						
Extensive	40	29.55	8.87	48.08	8.86	18.85
Regular	39	31.30	11.04	41.88	11.50	10.59
Jr college						
Extensive	31	16.74	8.00	33.71	9.02	17.06
Regular	18	17.56	7.42	25.69	10.15	7.50

Table 3. Judgments of summaries

		Good	Average or not good
Judge 1	pre	5	32
	post	13	24
Judge 2	pre	4	33
	post	22	15
Judge 3	pre	0	37
	post	7	30

Clearly, judges differed from each other in their ratings, but in all three cases, there was significant movement into the "good" category (for judge 1: chi square = 4.698, $df = 1$, $p < 0.05$; for judge 2: chi square = 19.212, $p < 0.01$; for judge 3: $p = 0.004$; Fisher Exact Probability Test: conditions for chi square test not met). When "good" and "average" categories are collapsed, results are similar.

Questionnaire. Students were asked if they felt their writing had improved. Thirty-six out of the 37 who answered this question thought it had, while one student was undecided. When asked if they thought reading helped them improve their writing, 32 out of the 38 who answered this question said "yes".

Discussion

The cloze test results of Experiment 2 replicate those of the first study. In addition,

the results on our measure of writing confirm that improvement in writing is possible without conscious learning; the extensive reading students had no special instruction in composition during the study, and there was almost no correction of grammar. In addition, most students felt they had improved in writing. Of course, it can also be argued that these gains were due to the actual writing students did (the "simple writing hypotheses"; the claim that writing without feedback will cause language acquisition), but this hypothesis has numerous difficulties, including findings that people, in general, write much too little to account for the complexity of the language they acquire, and studies showing no relationship between writing frequency and writing ability (research reviewed in Krashen, 1994). Nevertheless, Experiment 3 was designed to test this possibility.

Experiment 3: Extensive Reading With Writing In The Target Language

In the third study, groups differed in the language used for writing their summaries. In addition, a comparison group was used that had extensive work on cloze exercises, the format of the dependent variable. Another way in which this study differed from the others was the fact that more books were available. Experiment 3 was conducted at one of the authors' (B.M.) institutions, where a much larger choice was available, over 3000 books.

Procedure

Subjects. Three groups of subjects participated in this study.

1. English response group: students participated in an extensive reading program similar to that described in the previous two studies for one academic year. As in Experiment 2, students were required to write a short summary of the book. Group (1) wrote their responses in English.
2. Japanese response group: students did extensive reading, but wrote their summaries in their first language, Japanese.

3. Comparison group: the comparison group did extensive work on cloze exercises (approximately three to four hours per week, done in class with the help of the teacher and at home), read 32 stories, each four pages in length, and took a vocabulary test every week on words they had to look up to do the cloze exercises. One of the experimenters (B.M.) taught the class. To insure that the traditional/cloze method was given every chance to succeed, B.M. selected this particular class for use in the study because, of the three classes of this kind she taught, this one appeared to be the most proficient. B.M. also noted that they were an especially diligent group.

Both extensive reading groups did about three to four hours per week of homework. The experimental classes devoted this time to extensive reading, while the comparison group did cloze exercises and intensive reading with a dictionary.

Measures. The same cloze test that was used in Experiments 1 and 2 was given as a pre- and post-test. In addition, a test of reading comprehension was used, but only post-test scores were available for statistical analysis for the reading comprehension test.

The reading comprehension test was demanding: it consisted of about 20 pages of text (about 250-300 words per page) from an intermediate level (1600 word level) guided reader, which subjects were given one hour to read. There were 50 multiple choice questions, which subjects answered in 30 min. Reliability was estimated to be $r = 0.86$ (Kudar-Richardson formula, applied to comparison group).

In addition, subjects in all three groups were asked to write a summary in English of a book they read at the beginning of the school year, and were asked to repeat this task using the same book at the end of the academic year. The two summaries were evaluated holistically by two native speakers of English on a 1-6 scale, where 1 = worst and 6 = best writing. Specifically, raters were first asked to categorize papers as "good", "bad", or "average". Then papers in each pile were re-evaluated and divided into two groups to form six categories. As in Experiment 2, judges did not know whether the papers had been written at the beginning or end of the academic year.

Subjects were also asked how much time it took them to read the required book at the beginning of the year, and to indicate how much time it took them to read the same book at the end of the year.

Results

Cloze test. As shown in Table 4, both extensive reading groups made better gains than the comparison group on the cloze test ($F = 3.80$, $df = 2/111$, $p < 0.025$). The difference, however, was only significant for the English response group (protected $t = 2.74$, $df = 106$, $p < 0.01$); the difference between the Japanese response group and the comparison group was not statistically significant (protected $t = 1.09$, $df = 106$, ns). There was, however, no significant difference between the two extensive reading groups ($t = 1.071$). (The protected t-test, also known as the Fisher LSD test, is appropriate for multiple comparisons when the ANOVA F is significant and the total number of groups involved is not large; see Welkowitz *et al.*, 1982.)

Reading comprehension. On the reading comprehension test, both extensive reading groups were significantly better than the comparison group, and were not significantly different from each other ($F = 4.127$, $df = 2/106$, $p < 0.025$; English response vs comparison: $t = 2.09$, $df = 106$, $p < 0.05$; Japanese response vs comparison: $t = 2.72$, $df = 106$, $p < 0.01$; English response vs Japanese response: $t = 0.33$, ns).

Table 4. Extensive reading vs. traditional classes

Class	<i>n</i>	Cloze			Reading comprehension	
		Pre-test	Post-test	Gain	<i>n</i>	Post-test
Extensive: resp. in Japanese	40	29.45 (8.31)	45.52 (8.28)	16.08	36	70.50 (7.08)
Extensive: resp. in English	36	28.03 (9.02)	46.89 (6.58)	18.86	36	69.39 (7.62)
Comparison	38	30.13 (8.07)	44.29 (9.16)	14.16	37	65.57 (8.35)

Writing. The evaluations done by the two raters are presented separately, as it is possible that different criteria were used (Table 5).

Gains were significantly different for the three groups for each rater (rater 1: $F = 14.32$, $df = 2/99$, $p < 0.001$; rater 2: $F = 18.92$, $df = 2/99$, $p < 0.001$). Post hoc protected t-tests revealed that for both raters, the gains for those who wrote in Japanese were greater than the comparison group gains (rater 1: $t = 3.79$, $df = 66$, $p < 0.01$; rater 2: $t = 5.96$, $df = 66$, $p < 0.01$). Those who wrote in English actually gained less than comparisons according to the first rater, but the difference was not statistically significant ($t = 1.35$, $df = 66$, ns). According to the second rater, those who wrote in English made better gains than comparisons ($t = 1.90$, $df = 66$) which

came very close to the $p = 0.05$ level of significance for a two-tail test. Those who wrote in Japanese made significantly better gains than those who wrote in English (rater 1: $t = 5.14$, $df = 66$, $p < 0.01$; rater 2: $t = 4.09$, $df = 66$, $p < 0.01$).

Table 5. Rating of summaries written in English

	Pre-test	Post-test	Gain
Rater 1:			
Extensive reading: Japanese resp.	1.53 (0.79)	2.44 (1.05)	0.91
Extensive reading: English resp.	2.32 (1.15)	2.47 (0.896)	0.15
Comparison	2.85 (1.28)	3.21 (1.17)	0.35
Rater 2:			
Extensive reading: Japanese resp.	1.62 (0.985)	3.56 (1.13)	1.94
Extensive reading: English resp.	2.79 (1.15)	4.18 (1.47)	1.38
Comparison	2.91 (1.36)	4.03 (1.29)	1.12

Reading speed. Table 6 shows the amount of time subjects said they took to read the text at the beginning of the school year and at the end. The gains in reading speed were significantly different ($F = 79.80$, $df = 2/100$, $p < 0.001$). Protected t-test showed that those who responded in Japanese improved their speed significantly more than the comparison group ($t = 12.76$, $df = 68$, $p < 0.001$) and those who responded in English improved their speed more than the comparison group ($t = 5.70$, $df = 67$, $p < 0.001$). However, the extensive readers who responded in Japanese improved their speed significantly more than those who responded in English ($t = 6.77$, $df = 65$, $p < 0.001$).

Table 6. Time taken to read text (in min)

	<i>n</i>	Pre-test	Post-test	Difference
Extensive reading: Japanese resp.	34	40.12 (8.01)	18.32 (3.67)	21.80
Extensive reading: English resp.	33	30.73 (7.42)	18.21 (3.11)	12.52
Comparison	36	28.89 (5.31)	24.06 (5.76)	4.83

Pre: beginning of school year.

Post: end of school year.

Discussion

Experiment 3 was designed to tell us the effects of writing in the first language vs writing in the second language. One result supports the output hypothesis: those who wrote in the second language showed significantly better gains than comparison students on the cloze test, while those who wrote in their first language during the school year did not significantly outperform comparison students on the cloze test.

Other results, however, do not support the output hypothesis: the difference between the English response group and Japanese response group on cloze test gains was not statistically significant, and both groups of extensive readers did better than comparisons on the test of reading comprehension, but were not significantly different from each other. In writing, the Japanese response group clearly made better gains than both the English response and comparison groups, and they made the largest increase in self-reported reading speed.

It should be noted, however, that because the Japanese response group did not do as well on the first writing sample and read slower than the other groups, a floor effect might have contributed to their superior gains. (One means of controlling for this difference is the use of analysis of covariance. Use of ANCOVA with intact non-randomized groups is not considered appropriate, however; see Pedhazer, 1982.) It should also be noted that the Japanese response group caught up with the English response group and passed the comparison group on reading speed, which strongly suggests that their superior gains were real.

Summary

The clearest result of the series of studies presented here is that extensive reading was a consistent winner in all three studies. In the first study, "reluctant" EFL students did extensive reading for one semester and made superior gains compared to a traditional class, nearly catching up to them by the end of the semester. In the second study, which lasted for one year, extensive readers in four-year and two-year colleges outperformed traditional student. In the third study, extensive readers who wrote summaries of the books they read in English outperformed traditional students who focused on practicing cloze exercises on a test of reading comprehension and had better gains on a cloze test. A group that wrote their summaries in Japanese also outperformed comparisons, but the difference was only significant for the reading comprehension test. This group, however, made the best

gains in writing and reading speed, exceeding both the comparison group and the group that wrote summaries in English

The combined impact of the three studies can be presented in two ways. First, we can simply tally the number of cases in which the differences between the groups were statistically significant. Taking all cloze and reading comprehension comparisons into consideration, extensive readers made better gains in six out of seven cases and were worse in none. A more conservative version of this procedure is to only allow each group to count once. According to this procedure, extensive readers made better gains in four out of five comparisons (cloze test results only) and were worse in none.

A more precise way of summarizing the data is the use of effect size (Wolf, 1986; Light and Pillemer, 1984). While measures of statistical significance tell us the chances of obtaining a given result by chance, measures of effect size tell us the impact of the experimental treatment. Two common measures of effect size are the correlation coefficient (r) and " d ", where $d = (\text{mean of the experimental group} - \text{mean of control group}) / (\text{pooled standard deviation})$. (Some analysts use the standard deviation of the control group, rather than the pooled standard deviation; see Wolf, 1986, for discussion; d and r are related to each other by a simple algebraic formula; see Wolf, 1986: p. 35.)

Table 7. Effect sizes for the three experiments

Study	Subjects	Duration	Measure	Results	Effect size		
					d	r	n
1	Four-year college	One semester	Cloze	ER > Trad.	0.702	0.34	40
2a	Four-year college	One year	Cloze	ER > Trad.	1.11	0.49	79
2b	Two-year college	One year	Cloze	ER > Trad.	1.47	0.60	49
3	Four-year college	One year					
	(a) wrote resp. in Japanese		Cloze	ER = Cloze	0.244	0.12	78
			RC	ER > Cloze	0.609	0.30	73
	(b) wrote resp. in English		Cloze	ER > Cloze	0.630	0.30	74
			RC	ER > Cloze	0.480	0.24	73

ER = extensive reading;

Cloze = traditional instruction with emphasis on cloze exercises.

Effect sizes were calculated using Johnson (1989), and were calculated from t-scores and sample sizes, a procedure equivalent to the formula presented above. Table 7 presents cloze and reading comprehension test effect sizes for the individual studies. All effect sizes are positive, confirming that extensive readers outperformed comparison students in all comparisons. The mean effect size for the entire sample was 0.749; this was recalculated, taking sample sizes in account, resulting in an adjusted mean effect size of $d = 0.730$ (see Wolf, 1986; p. 41 for procedures). This means that the average student in the extensive reading treatment scored 0.73 standard deviations above the average student in the comparison group.

Similar calculations were carried out for only those studies using the cloze test as a dependent variable, which meant that each group was counted only once, and only gain scores were utilized in effect size calculation. The mean effect size was 0.831, and the adjusted mean effect size was $d = 0.813$.

Inspection of the effect sizes in Table 7, however, reveals clear variability. This was confirmed by application of the Test of Homogeneity, described in Wolf (1986). For both the entire sample as well as just those comparisons utilizing cloze tests, there was significant heterogeneity (full sample: chi square = 17.10, $df = 6$, $p < 0.001$; reduced sample: chi square = 15.35, $df = 4$, $p < 0.01$). The causes of this heterogeneity are not fully clear: one could hypothesize that the large effect size in the second study was due to the longer treatment; this explains why the effect sizes in the second study are larger than the first, but does not explain why the effect sizes in the third study are not larger, especially in view of the fact that a wider selection was available to the students. It is, however, interesting that the comparison group in Experiment 3 made better gains than did the comparison group in Experiment 2, suggesting that they did more reading or that the cloze exercises contributed some comprehensible input.

Despite this variation, the results are remarkably consistent. The reliability of the advantage of the extensive readers, as well as the results of previous studies, also helps reduce the potential harm caused by the fact that intact classes were used; it was not possible to randomize subjects. Repeated experimentation with different classes, however, provides quasi-randomization.

Accountability

In most sustained silent reading (SSR) studies reported in the research literature, no accountability was used; subjects simply read for pleasure and were not required to do any sort of book report. In the series of studies reported here, there was some

accountability, but it was not excessive: readers were required to read, to report on what they read, and had incentives (course credit and grades).

Our results thus suggest that the positive effects of SSR can be maintained even when some of the conditions are slightly weakened. There appears, however, to be a limit on just how far one can go. Carver and Liebert (1995) reported no gain in vocabulary knowledge for elementary school students after a six-week (24 hr) extensive reading program. Carver and Liebert clearly departed from SSR too much; there was heavy use of extrinsic motivators, their subjects were restricted to "easy" books, at or below their reading level, and had to take multiple choice test on the books they read. In addition, reading time was heavily concentrated, with students reading in two-hour blocks, with breaks, a procedure that also departs from usual SSR practice.

Our data do not provide us with any way of determining whether accountability was necessary for the success of extensive reading, but it is clear from other studies that free reading can work without it. It remains to be determined (1) whether the "light" accountability used in this study was helpful, (2) if it was helpful, what level of accountability is optimal, and (3) whether accountability is more effective in some situations than others (e.g. EFL vs ESL; children vs adults).

The role of output

Study 3 attempted to determine whether the fact that students wrote their summaries in English had any impact on their progress. While the results of one comparison supported the output hypothesis, most of the data did not. In fact, those who wrote summaries in their first language made superior gains in writing and reading speed.

Conclusion

Overall, extensive reading proved to be superior to traditional approaches on measures of reading comprehension, as well as on measures of writing and reading speed, and, according to teacher observations, was much more popular with students. Perhaps the most telling observation in support of the latter claim is the reaction of students in the third comparison group, the one that focused on cloze exercises. These students used the same classroom used by one of the extensive reading classes, and saw the books displayed and other students checking out

books. This led to frustration: they asked for books and felt cheated that their classwork did not include extensive reading.

We presented here the results of three studies, all using an intact class quasi-experimental design. Extensive reading resulted in significantly superior gains in six out of seven comparisons for cloze and reading comprehension tests and extensive readers did better on measures of writing and reading speed. Extensive reading allowed "reluctant" students of EFL to catch up to traditional students, and worked in a variety of EFL situations. When we add these results to those already published in second language acquisition, and to the massive support free reading has in first language development, the case for self-selected reading for pleasure is overwhelming.

References

Carver, R. and Lieber, R. (1995) The effect of reading library books at different levels of difficulty upon gain in reading ability. *Reading Research Quarterly*, **30**, 26-48.

Cho, K. S. and Krashen, S. (1994) Acquisition of vocabulary from the Sweet Valley Kids series: adult ESL acquisition. *Journal of Reading*. **37**, 662-667.

Cho, K. S. and Krashen, S. (1995a) Becoming a dragon: progress in English as a second language through narrow free voluntary reading. *California Reader*, **29**, 9-10.

Cho, K. S. and Krashen, S. (1995b) From Sweet Valley Kids to Harlequins in one year: a case study of an adult second language acquirer. *California English*, **1**(1), 18-19.

Constantino, R. (1994) Pleasure reading helps, even if students don't believe it. *Journal of Reading*, **37**, 504-505.

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

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

Hafiz, F. and Tudor, I. (1989) Extensive reading and the development of language skills. *English Language Teaching Journal*, **43**, 4-13.

Johnson, B. (1989) *DSTAT: Software for the Meta-Analytic Review of Research Literature*. Erlbaum, Hillsdale, NJ.

Krashen, S. (1994) The input hypothesis and its rivals. In *Implicit and Explicit Learning of Languages*, ed. N. Ellis, pp. 45-77. Academic Press, London.

Light, R. and Pillemer, D. (1984) *Summing Up: The Science of Reviewing Research*. Harvard University Press, Cambridge.

Pedhazet, E. (1982) *Multiple Regression in Behavioral Research*, 2nd edn. Holt Rinehart Winston, New York.

Pilgreen, J. and Krashen, S. (1993) Sustained silent reading with English as a second language high school students: impact on reading comprehension, reading frequency, and reading enjoyment. *School Library Media Quarterly*, **22**, 21-23.

Tudor, I. and Hafiz, F. (1989) Extensive reading as a means of input to L2 learning. *Journal of Research in Reading*, **12**, 164-178.

Welkowitz, J., Ewen, R. and Cohen, J. (1982) *Introductory Statistics for the Behavioral Sciences*, 3rd edn. Academic Press, New York.

Wolf, F. (1986) *Meta-Analysis: Quantitative Methods for Research Synthesis*. Sage, Newbury Park, CA.