Indonesian Journal of English Language Teaching
Volume 5/Number 1 • May 2009

A THREE-YEAR LONGITUDINAL STUDY OF IN-CLASS SUSTAINED SILENT READING WITH TAIWANESE VOCATIONAL COLLEGE STUDENTS

Ying-ying Hsu
Nanya Institute of Technology

Sy-ying Lee
National Taipei University

Abstract

This study examined the effects of three years of in-class sustained silent reading with a group of vocational college students in Taiwan. Readers outperformed comparisons on tests administered after one semester and increased their advantage on tests given at the end of the first year. The gap between the groups narrowed the second year, but readers maintained their superiority at the end of the second and third year. The initial gains were probably due, in part, to the Hawthorne Effect. It is likely that the progress made in the second and third year were more modest, because of external demands on students' time, which limited the amount of reading students could do outside of class. Overall, the results clearly support the practice of in-school self-selected reading.

Keywords: In-class sustained silent reading; vocational college student; Hawthorne Effect; in-school self-selected reading

INTRODUCTION

It has been widely accepted that extensive reading or free reading is essential for language and literacy development (Goodman, 1982; Krashen, 1982; Smith, 1983).

Despite some variation due to slightly different approaches or the duration of the study, most studies have reached a strikingly

Direct all correspondence to:

Ying-ying Hsu, Department of Applied Foreign Languages, Nanya Institute of Technology, No. 414, Sec. 3, Jhongshan E. Rd., Jhongli City, Taoyuan County 32091, Taiwan. Email: yhsu@yahoo.com

Sy-ying Lee, Department of Foreign Languages and Applied Linguistics, National Taipei University, 151 Da Xue Rd, Sanxia Town, Taipei County 23741, Taiwan. Email: syying.lee@gmail.com

similar conclusion: both sustained silent reading integrated into a curriculum and after-school extensive free reading help with English acquisition in vocabulary, grammar, reading comprehension, writing and listening comprehension (for further review, see Krashen 2004, Lee 2005a, 2007, & Mason 2003).

Lee (2007) has concluded that at least three conditions need to be met to make a reading program successful: **sufficient book access**, **sufficient program duration** (for at least one year), and **full self-selection** with help from the teacher when needed. These conditions correspond perfectly with the Comprehension Hypothesis (Krashen, 1982) which states that we acquire language through obtaining a large quantity of input that is comprehensible and compelling to us.

In light of the previous success of SSR among university students in Taiwan (Lee, 2007; Liu, 2007; Sheu, 2004; Smith, 2007), it was of interest to us to see if SSR could have the same impact on vocational college students who were somewhat lower in academic performance than university students, and who probably also have less competence in English.

In a pilot study with vocational college students (the same subjects as those in the present study; Hsu & Lee, 2005), those who scored higher on reading and vocabulary pre-tests were also those who had had the experience of free reading in English when younger, a finding that supports the hypothesis that we learn to read by reading (Goodman, 1982). Those who read more also did better on the pre-test essay on writing even before they began receiving formal writing instruction in school, confirming the finding that free reading is the single most important variable predicting writing performance (Lee, 2005b).

In this study, we investigate the effects of extensive reading over a long term, integrating extensive reading into a curriculum for three years for vocational college students, by far the longest longitudinal study of this kind attempted with second language acquirers, a severe test of the efficacy of extensive reading.

METHOD

Subjects, Instructors and Classes

At the onset of the project, 86 3rd year junior college students (aged 17 to 18) in the Department of Applied Foreign Languages at a vocational junior college participated in this project, 43 in the experimental group (the SSR group) and 43 in the comparison group.

They had studied English for five years, but had had no formal English writing instruction in the first two years of junior college, or in any previous schooling. The numbers of subjects taking the planned measures, however, changed with time, with some students transferring, some dropping out, and some absent when the tests were administered.

The first author was the instructor for the experimental group for both reading and writing for the first two years of the project, as well as the comparison group for the first year. The second year, another teacher took over the comparison group writing class. After the second year, reading class was not continued, and students had courses in other areas, such as language and culture, second language acquisition, and English for Special Purposes. The comparison group did not do any SSR in their classes.

In the first year, SSR was done as part of the reading class. Both groups received the same instruction in writing, using the same text and all writing about three to four essays in a semester. In the second year, SSR was done as part of the writing class. As a result, the SSR group students received less formal instruction and wrote fewer essays than the comparison group did.

The Procedure: The Three-year Plan

Appendix A shows the whole picture of how this longitudinal study proceeded with use of different sets of measures at different stages during the three years. Appendix B presents the study with the language aspects measured: vocabulary, reading, and writing.

It may have been preferable if we had used the same measures throughout the study. The use of the same measures, however, would have caused some severe methodological problems, as students may have gotten familiar with the tests. One of the solutions to this problem is to have different but equivalent forms for each of the measures that have been tested for their validity and reliability. Such measures, however, were not available, and it was not financially possible for students to take nationwide or international standardized tests, such as TOEFL, regularly throughout their school years.

Effect Sizes and the Measures Used

The use of effect sizes helps us cope with the problems mentioned above. Effect sizes are a measure of the strength or impact of a treatment and allow comparisons when different (but reasonably similar) measures are used. We used effect sizes and T values to present

our results so that both the differences between groups and the patterns of growth could be observed.

All measures used in this project have been tested for their validity and reliability and have been widely used in other studies, such as Nation's (1990) and Schmitt's (2000) vocabulary level tests. Huang's tests (2003) are popular among researchers in Taiwan studying vocational students' language proficiency and problems. Cloze test A was validated and used in Mason (2003). Cloze B, also created by Mason, has a test-retest reliability of .87 (Cronbach's alpha) (See Hsu & Lee, 2007).

Four essays were collected at different stages: (1) pre-essays were collected before the project began, (2) post-essay 1 was collected at the end of the first year, (3) post-essay 2 in March of year 2 (two months before the academic year ended), and (4) post-essay 3 was collected at the end of year three. 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), language use (5-25), and overall impression (34-100).

For the overall impression rating, Jacob et al. (1981) placed different weightings on each subscale, with more value placed on content and structure, and less on mechanics. The total number of words written was also counted as a measure of fluency. Finally, a reflection paper was collected probing how students reacted to this long-term project.

The Grading of the Essays

To ensure sufficient inter-rater reliability, three raters were involved in the grading task. All were experienced raters, university professors of English who had served as raters for the nationwide entrance examination for over 10 years.

Two raters did the initial grading. When there were gaps over 10 points, one tenth of the total score, a third rater took the position to assign a final score. The essays of the two groups were mixed before they were handed to the raters, so the raters would not be influenced when scoring. Kendall's W was employed to analyzed inter-rater reliability. In each set of the collected essays, fewer than 10 essays had scoring gaps of over 10 points. The Kendall's Ws were .87, .82, .80, and .94 for the four essays respectively, all significant at the .01 level.

The Reading Materials

Increasing numbers of graded readers were brought into the class at each stage, with 354 *Penguin* readers for the pilot study. One hundred seventy-six more readers from the *Oxford* series were added for the first year study. For year 3, we added some authentic reading materials such as *Goosebumps, The Magic Tree House, The Secret of Droon,* and some books from the *June B. Jones* and *Sweet Valley* series. By year 3, approximately 700 titles were circulating in the classroom, but most students had stopped reading the very low level readers after the second year. The SSR group was required to complete a very simple reading log each week, which could be written in either language. The researchers decided to drop this requirement in year 3, because of students' fuller schedule before graduation. Some enthusiastic students still kept doing the log after it became optional.

In the following section, we present the pattern of development in vocabulary, reading, and writing after three years of experiencing in-class SSR.

RESULTS

This section reports the results by providing the raw scores and the effect sizes along with a multivariate analysis—Hotelling's T-Square—that allows comparisons of mean values of two groups with more than one pair of dependent variables. Cohen (1988) recommends the following benchmarks for interpreting effect sizes: .2 = small effect; .5 = moderate effect; .8 or larger = large effect.

Starting from the first stage of the study, indicated by the T values, no significant differences were found between the two groups on pre-tests; both groups also performed similarly after one semester, although the experimental group was slightly better on vocabulary and Cloze A (Table, 1 & 2).

TABLE 1
Vocabulary and Cloze Raw Scores for Both Groups

	Pre	one semester	one year	2 years	3 years
	(Nation,1990)		(Huang, 2003)	(Schmitt, 2000)	
1000- n[a]			41, 41		
EXP			12.61 (2.49)		
COMP			12.49 (2.88)		
2000 - n	43, 43	43, 43	41, 41	42, 37	41, 41
EXP	11.8 (5.33)	14.05 (4.52)	9.46 (3.84)	18.43 (6.0)	22.61 (4.75)
COMP	10.28 (4.86)	13.28 (4.36)	8.07 (3.39)	17.65 (5.30)	19.24 (7.05)
3000 - n	43, 43	43, 43	41, 41	42, 37	41, 41
EXP	6.58 (4.32)	8.65 (4.64)	8.51 (3.38)	12.21 (6.25)	15.15 (6.83)
COMP	6.44 (2.88)	7.6 (3.59)	6.71 (3.35)	10.35 (4.98)	13.24 (6.66)
5000 – n				42, 37	41, 41
EXP				9.71 (5.33)	12.66 (6.08)
COMP				8.35 (3.79)	10.46 (6.06)
	Pre	**one semester**	**one year**	**2 years**	**3 years**
	Cloze A (Mason, 2003)		**Cloze B (by Mason[b])**		
Cloze- n	43, 43	43, 43	41, 41		
EXP	26.84(12.67)	32.28 (13.49)	29.35(10.52)		
COMP	24.86(11.26)	27.47 (11.63)	21.66(10.16)		

a. n = the number of subjects taking the test
b. Reliability of Cloze B = .87 (Cronbach's alpha), with vocational college students (Hsu & Lee, 2007).

TABLE 2
T values & Effect Sizes for the Vocabulary Level Tests and Cloze Tests

	Pre		one semester		one year		2 years		3 years	
	Hotelling's T= .04 p=.22		Hotelling's T= .02 p=.51		Hotelling's T= .18 p =. 01*		Hotelling's T= .04 p =.42		Hotelling's T= .11 p =. 04*	
	t	**ES**	**t**	**ES**	t	**ES**	t	**ES**	t	**ES**
1000					.21	0.05				
2000	1.37	0.3	.80	0.12	1.74	0.39	.61	0.14	2.48*	0.57
3000	.18	0.04	1.17	0.26	2.43*	0.54	1.45	0.33	1.29	0.29
5000							1.30	0.29	1.76	0.37
cloze	.76	0.17	1.78	0.39	3.35*	0.75				

It can be seen that effect sizes for vocabulary rise nicely to the end of year 1, then drop in year 2 to the first semester level. The 3,000 VLT score holds steady, and the 2,000 VLT increases. No differences were found for the 1,000 and 5,000 VLTs.

Effect sizes for cloze tests progressed with large leaps to the end of year 1. We stopped giving cloze tests for the last two years because each test contained 100 items which had tired the students greatly. We thus focused on examining the effect of reading using only the vocabulary and writing tests in the second and third years.

TABLE 3
Raw Scores on Writing for Both Groups

	Pre N = 43, 43	**Yr 1 (post 1) N = 41, 40**	**Yr 2 (post 2) N = 43, 35**	**Yr 3 (post 3) N = 43, 36**
Fluency				
EXP	98.47 (34.86)	123.02 (37.06)	127.47 (36.49)	299.92 (82.32)
COMP	97.76 (30.14)	82.17 (34.86)	76.94 (17.55)	198.67 (51.77)
Content				
EXP	16.65 (2.39)	19.34 (2.46)	21.33 (3.85)	21.55 (3.7)
COMP	16.65 (2.36)	16.88 (2.73)	19.97 (3.2)	19.0 (2.46)
Organization				
EXP	9.79 (1.52)	12.79 (1.87)	14.21 (1.99)	14.6 (2.17)
COMP	10.04 (1.52)	10.54 (2.09)	13.54 (2.02)	12.86 (2.1)
Vocabulary				
EXP	9.76 (1.52)	11.62 (1.44)	14.21 (1.99)	14.05 (2.15)
COMP	10.23 (1.46)	10.10 (1.74)	12.89 (1.83)	12.78 (1.62)
Language use				
EXP	9.76 (2.63)	12.73 (2.51)	15.67 (3.24)	15.74 (3.48)
COMP	10.65 (2.28)	10.78 (2.52)	13.71 (2.53)	14.21 (2.3)
Mechanics				
EXP	2.99 (.54)	3.33 (.61)	3.86 (.87)	3.38 (.87)
COMP	3.12 (.69)	3.02 (.61)	3.14 (.6)	2.94 (.62)
Impression				
EXP	48.95 (8.01)	59.82 (8.01)	69.26 (10.93)	69.29 (11.68)
COMP	50.69 (7.68)	51.32 (7.68)	59.84 (16.98)	62.04 (7.84)

TABLE 4
T Values and Effect Sizes for All the Aspects of the Writing Measurement

	Pre		Yr 1		Yr 2		Yr 3	
	Hotelling's T= .12, p = .16		Hotelling's T= .62, P<.000		Hotelling's T= .90, p <.000		Hotelling's T= .58, p = <.000	
	t	ES	t	ES	t	ES	t	ES
Fluency	.09	0.02	5.49*	1.23	7.42*	1.73	6.39*	2.7
Content	.01	0	4.29*	0.96	1.64	0.39	3.55*	0.81
Organization	-.74	-0.17	5.16*	1.15	1.32	0.29	3.59*	0.82
Vocabulary	-1.45	-0.32	4.32*	0.96	2.93*	0.7	2.91*	0.67
Use	-1.68	-0.37	3.52*	0.78	2.94*	0.67	2.27*	0.54
Mechanics	-.98	-0.21	2.26*	0.51	3.92*	0.93	2.53*	0.58
Impression	-1.03	-0.22	4.37*	1.1	2.58*	0.68	3.17*	0.73

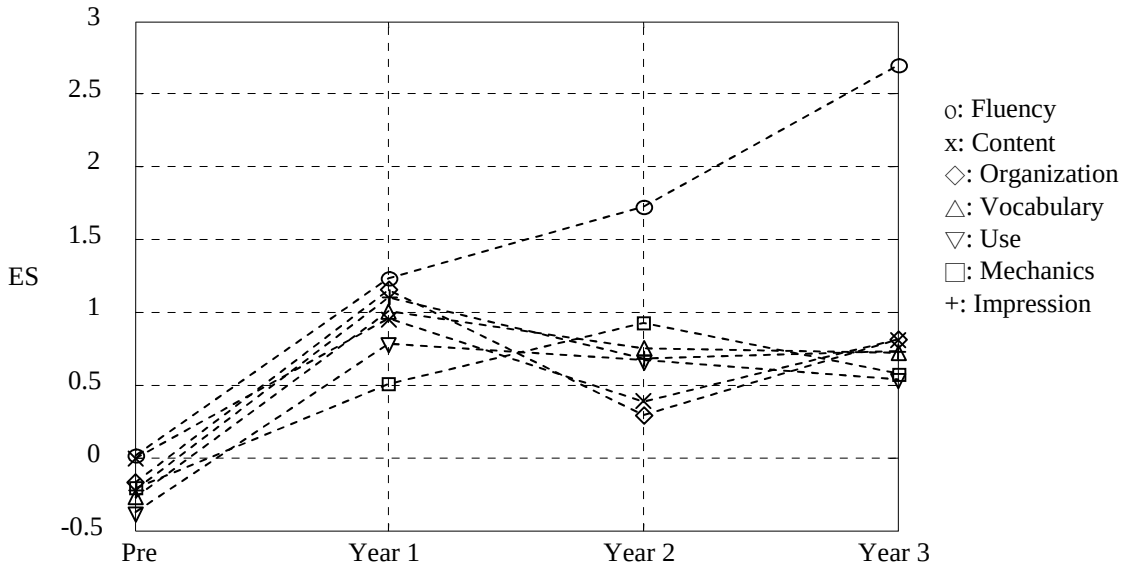

Figure 1 The Growth Pattern of Writing for the Experimental Group

Tables 3 and 4 present results of the writing test, and Figure 1 presents the growth pattern of the experimental group in terms of effect sizes, that is, in terms of its advantage over the comparison group.

The SSR group performed consistently better than the comparison students in writing (Table 3). The negative signs in Table 4 mean that the experimental students lost to the comparison group on the pre-tests. Note, however, that they made rapid improvement and outperformed the comparison students in the first year. As shown in Figure 1, the

experimental group lost some of its advantage during the second year, except for fluency and mechanics, but on the average they were still better than their counterparts.

In year 3, the experimental group regressed a little on vocabulary, language use, and mechanics, but recovered well in content, organization and overall quality. They were well ahead of comparisons in fluency. The overall gain for three years is still large, as suggested by the effect sizes at the end of year 3.

DISCUSSION

This three-year study began as a one-semester pilot study (Hsu & Lee, 2005), which lead to full academic year study (Lee & Hsu, 2007; Hsu & Lee, 2007), and eventually culminated after three years when the subjects graduated from the vocational junior college.

The experimental group steadily increased its advantage over the comparison group during the first year on vocabulary and cloze tests. The cloze was not given after year 1. After year 1, the gap between the two groups narrowed on vocabulary tests, except for the 2,000 VLT in year 3, where the experimental group increased its advantage over the comparison group. It was no surprise that no differences were found for the 1,000 and 5,000 VLTs. Taiwanese scholars have found that the average vocabulary size for a vocational high school student is approximately 1,690 words (Huang, 2003), and in 1999 it was reported that university students may know about 3,500 English words (Chen, 1999), which may be higher now. The average vocational college student's progress may thus be limited to the 2,000 and 3,000 word levels.

The results for writing were much stronger in favor of the SSR group, but there was a similar drop in the experimental group advantage after year 1. After three years, however, the experimental group was still superior to the comparison group.

The rise in scores in the first year may have been due to the Hawthorne Effect, the enthusiasm students experience for a new approach. The subsequent decline has a clear explanation as well: the first author observed that most students in the reading group read less and less in the second and third years of the project. Nearly all the experimental students described in their reflection papers how busy they became in their fourth and fifth years of school. The most frequently mentioned factors preventing them from reading more were working part time, retaking courses, and attending cram schools to prepare for examinations (such as entrance examinations for university or

professional license tests). Students were able to devote time to reading outside of class in year 1, but it became more difficult to find the time during the next two years.

An additional factor affecting the writing scores is the fact that for the first three essays, students were given descriptive topics, such as "My best friend," "My favorite season," and "The Moon Festival." For the third year, the essay was argumentative, on the question of whether students should or should not work part time. Both groups were informed of the kind of essay they were about to write and were encouraged to do some brainstorming at home before they came to class.

The SSR group performed only slightly better than the comparison group on this topic, and the difference for content and organization was not statistically significant.

It must be pointed out that comparisons had received a great deal of instruction in their classes on strategies used to write argumentative essays. Thus, SSR students did a little better those who had received more training on strategies of argumentative writing.

If these explanations are correct, they suggest that it is important to keep SSR time in class, as it is difficult to make reading a regular behavior because of conditions not under students' control. Pressure from schoolwork, part time jobs, and examinations in and out of school may have diminished students' desire or willingness to do more reading, especially when other homework appeared to be more directly related to ongoing classroom activities and tests.

For this reason, an in-class SSR time slot is a necessity. In addition, offering reading materials carefully selected for this group of students, including graded readers and authentic materials for teens, means that students do not have to worry about searching out and buying books on their own. This in-class SSR practice was the only variable that made the experimental students better in writing throughout the longitudinal study.

Although reading was not as enthusiastically done toward the end of the project, there was a steady positive effect of reading on students' English writing, as seen in Figure 1. In fact, the early rise and subsequent mild decline followed by a recovery may be just what we should expect for the practice of in-class SSR. To see if reading results in consistent gains, we may need to observe for an even longer duration; Figure 1 may have only captured a small segment of an uphill zigzag of development.

Two additional issues are of concern: (1) the input from other courses related to and taught in the target language, and (2) the lack of

appropriate measurement tools developed to examine gains or improvement in different stages. The former problem is dealt with by having a comparison group and making sure that the in-class SSR was kept in either the reading class or the writing class and was done only with the experimental group. As for the second problem, using effect sizes can help control for the pre-test difference, allows us compare test results even when somewhat different measures are used, and can help us examine the growth pattern when a new treatment is applied.

Our results do not show that reading alone can do all the work in developing literacy in a second language; it is likely that our subjects also acquired a great deal from other sources of comprehensible input. Our results do show, however, that regular self-selected reading has once again proved itself to be crucial and fundamental practice our students need.

THE AUTHORS

Ying-ying Hsu, MA in TESOL, Senior Lecturer of English at Nanya Institute of Technology, Taiwan. She collaborated with Sy-ying Lee in a couple of projects tracing how in-class sustained silent reading effects English acquisition and has had several publications coauthored with Prof. Lee.

Sy-ying Lee, Ph.D., Professor, Department of Foreign Language and Applied Linguistics, National Taipei University, 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.

REFERENCES

Chen, H. J. (1999). How many words do they know? Assessing Taiwanese college EFL students' receptive and productive vocabularies. *Proceedings of the 16th Conference on English Teaching and Learning in the Republic of China*.

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

Goodman, K. (1982). 'Language, literacy and learning'. *Language and literacy: The selected writings of Kenneth S. Goodman.* London: Routledge.

Hsu, Y. Y., & Lee, S. Y. (2005). Does extensive reading also benefit junior college students in vocabulary acquisition and reading ability? *Proceedings of the 22nd International Conference in English Teaching and Learning.*

Hsu, Y. Y., & Lee, S. Y. (2007). Extensive reading and EFL junior college in Taiwan. *Studies of English Language and Literature, 20,* 137-145.

Huang, C. C. (2003). An investigation into vocational senior high school students'vocabulary knowledge and reading comprehension. *Studies in English Language and Literature, 12*, 39-58.

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.

Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Prentice Hall.

Krashen, S. (2004). *The Power of Reading*. (2nd ed.). Portsmouth, NH: Heinemann.

Lee, S. Y. (2005a). How robust is in-class sustained silent reading? *Studies in English Language and Literature, 15*, 65-76.

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

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

Lee, S. Y., & Hsu, Y. Y. (2007). Does extensive reading work with vocational college students? *Proceedings of 2007 International Conference and Workshop on TEFL & Applied Linguistics*.

Liu, C. K. (2007). What role self-selected reading can play in Freshman English Classes. *Proceedings of 2007 International Conference and Workshop on TEFL & Applied Linguistics*.

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

Mason, B. (2005) Vocabulary Acquisition through Storytelling. TexTESOL III Newsletter. February, 3-5.

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

Schmitt, N. (2000). *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

Sheu, P. H. (2004). The effect of extensive reading on learner's reading ability development. *Journal of National Taipei Teachers College, 17* (2), 213-228.

Smith, F. (1983). *Essays into Literacy*. Portsmouth, NH: Heinemann.

Smith, K. (2007). A comparison of "pure" extensive reading with intensive reading and extensive reading with supplementary activities. *Proceedings of 2007 International Conference and Workshop on TEFL & Applied Linguistics*.

APPENDICES

APPENDIX A

Stages and Dates of Data Collection

	Stages	Dates of data collection	Measures	Terms			
1st year	1	Sep. 2004	Pre-voc 2,000, 3,000 (Nation); Cloze-A (Mason); Pre-essay	A one-semester study (Hsu & Lee, 2005)	A one-year study (Lee & Hsu, 2007; Hsu & Lee, 2007)	A two-year study	A three-year study
		Jan. 2005	Post-voc 2,000, 3,000 (Nation); Post-Mason Cloze-A				
	2	June 2005	Voc 1,000, 2,000, 3,000 (Huang); Cloze-B (Mason); Post-essay 1				
2nd year	3	Sep. 2005	No data collected				
		Jan. 2006	Voc 2,000, 3,000, 5,000 (Schmitt)				
		March 2006	Post-essay 2				
3rd year	4	Sep. 2006	No data collected				
		May 2007	Voc 2,000, 3,000, 5,000 (Schmitt) Post-essay 3				

Direct all correspondence to:

Ying-ying Hsu, Department of Applied Foreign Languages, Nanya Institute of Technology, No. 414, Sec. 3, Jhongshan E. Rd., Jhongli City, Taoyuan County 32091, Taiwan. Email: yhsu@yahoo.com

Sy-ying Lee, Department of Foreign Languages and Applied Linguistics, National Taipei University, 151 Da Xue Rd, Sanxia Town, Taipei County 23741, Taiwan. Email: syying.lee@gmail.com

APPENDIX B

Measures Used At Different States of the Study

	The first semester (Sep. 2004 – Jan. 2005)	**The first year (Jun. 2005, at the end of the first year)**	**The second year (March 2006, in the middle of the second semester)**	**The third year (May 2007, at the end of the third year)**
Vocabulary	Pre-test and post-test using Nation's 2,000 VLT & 3,000 VLT	Huang's 1,000, 2,000, & 3,000 VLTs	Schmitt's 2,000, 3,000 & 5,000 VLTs	Schmitt's 2,000, 3,000 & 5,000 VLTs
Reading	Pre-test and post-test using Mason's Cloze A	Mason's Cloze B		
Writing	Pre-essay	Post-essay 1	Post-essay 2	Post-essay 3