Improving EFL learners' writing through enhanced extensive reading

Aaron David Mermelstein
Soochow University
Taipei, Taiwan

Abstract

This study's purpose is to seek out methods of improving reading and writing for EFL learners. This one-year study focuses on an enhanced design of extensive reading (ER) towards improving learners' writing abilities. Pre- and posttests used the Jacobs, Zingraf, Wormoth, Hartfield, and Hughey (1981) measurement of writing, including content, organization, vocabulary, language use, and mechanics. A sixth subscale, fluency, was also added. The results indicate significant differences in gains on all of the subscales favoring the treatment group. A measurement of effect size also demonstrated small to large effects across the six subscales. This study demonstrates that an enhancement of previously established ER protocols can achieve significant gains and sizable effects among learners.

Keywords: extensive reading, writing, English as a foreign language

Language teachers throughout the world are continually looking for methods of improving their students' language abilities. Time is always a key factor in both teacher lesson planning and learner language acquisition because there never seems to be as much of it as thought necessary. Teachers must teach curriculum and other requirements and often have students at mixed ability levels in their classrooms. Students have to deal with the workload and pressures of being a student, which naturally means learning more than one subject at a time and preparing for multiple tasks. Adult learners often have even more life pressures. Considering all of this, it is necessary for teachers and researchers to look for methods of enhancing or streamlining the learning process to be more time effective when teaching language skills.

Several studies have confirmed that reading more is connected to better writing skills in both first and second language (Applebee, Langer, & Mullis, 1986; Huang, 1996; Janopoulos, 1986; Lee, 2001, 2005; Lee & Hsu, 2009; Lee & Krashen, 1996, 2002). Specifically, extensive reading (ER) has been widely advocated for language learning (e.g., Beglar & Hunt, 2014; Day & Bamford, 2002; Waring, 2006; Yamashita, 2013). Susser and Robb (1990) defined ER as: (a) reading large quantities of material or long texts for global or general understanding with the intention of obtaining pleasure from the texts, (b) individualized reading with students selecting the texts they want to read, and (c) not being required to discuss the book in class. In this study, ER is defined as reading as much as possible within the learner's peak acquisition zone, for the purpose of gaining reading experience and general language skills.

Previous Studies Related to ER and Writing Improvement

Over the past few decades, several classroom studies have been carried out to determine if adding ER into the classroom can effectively improve the writing abilities of EFL learners (e.g., Elley, 1991; Elley & Mangubhai, 1983; Hafiz & Tudor, 1989, 1990; Lai, 1993; Lee & Hsu, 2009; Mason & Krashen, 1997; Tsang, 1996). Five of these studies (Hafiz & Tudor, 1989, 1990; Lai, 1993; Lee & Hsu, 2009; Tsang, 1996) specifically examined the impact of a reading program on the writing abilities of EFL learners coming from a similar educational background. All five of these studies also used descriptive writing as an evaluation of the participants' ability, and the duration of the reading programs lasted from 4 weeks to 30 weeks.

In general, all five of these studies reported significant results with the participants performing at higher levels than their comparison groups in several key areas. The most reported improvement came from Lee and Hsu (2009), who reported significant gains in five key areas: fluency, content, organization, language use, vocabulary, and mechanics. Hafiz and Tudor (1989, 1990) and Tsang (1996) all reported that their treatment groups had higher measurements of language use, specifically in syntax and semantics. Hafiz and Tudor (1989, 1990) and Lai (1993) reported that their treatment groups had increased ability in vocabulary use and fluency. Hafiz and Tudor (1990) also reported more variety in the vocabulary used by the participants. Tsang (1996) also reported that the participants improved in the content of their writing. However, not all of these studies had positive findings.

Some of the previously mentioned studies had negative findings. Tsang (1996) found that the treatment group did not outperform the comparison group in the areas of spelling, vocabulary, and organization. Hafiz and Tudor (1990) also reported that their treatment group did not significantly outperform their comparison group in the area of vocabulary. When Lee and Hsu (2009) compared their study with the other studies noted above, they suggested that the reason for the lack of significant improvement in some of the key areas was based on the design flaws of the other studies. They highlighted the point that all four of the previous studies only allowed the participants a limited selection or a small amount of reading materials. They also pointed out that in all four other studies, the participants had been required to make either written or oral reports regarding their readings, and this could have "extinguished some of the pleasure of reading" (p. 13). The third flaw they highlighted was the duration of the studies, with two of the shorter studies lasting only four weeks (Lai, 1993) and three months (Hafiz & Tudor, 1990).

It appears that the design flaws of these other studies noted by Lee and Hsu (2009) had a large influence over their own study's hypothesis and design. Lee and Hsu's hypothesis was that "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" (p. 13). Therefore, their study lasted for 30 weeks and required their participants to read for 50 minutes in class every week. They also offered their participants over 500 graded reader books and suggested that each participant read at least one book per week. In addition, Lee and Hsu only asked their participants to fill out a reading log and to write a brief reflection paragraph or summary upon completion of reading each book, believing this to be less accountability on the part of the

students than in the other studies noted above.

As previously mentioned, the reported results from Lee and Hsu (2009) noted significant gains in six key areas: fluency, content, organization, language use, vocabulary, and mechanics. These results were based on the criteria established by Jacobs, Zinkgraf, Wormuth, Hartfield, and Hughey (1981), plus their addition of fluency as a subscale. However, while these results are promising, there are three concerns with the design of the Lee and Hsu (2009) study. The first concern is regarding the practicality of their ER treatment, which consisted of 50 minutes of reading each week, or an entire class period. In this researcher's opinion, 50 minutes of uninterrupted reading is too long of an activity for most second language (L2) learners, let alone any learner. Not only is the length of this activity challenging for the learners, but it may also be impractical for most EFL teachers to devote so much classroom time each week. Teachers who are already overburdened with too much material to cover may be reluctant to devote any time to an ER activity (Takase, 2002). Many EFL teachers may not be in a position to devote 33% or more of their classroom time to reading due to the curriculum and/or other restraints (Helgesen, 2005), as they did in the Lee and Hsu (2009) study.

The second concern is based on the ER treatment. Lee and Hsu (2009) stated that the "experimental group students were allowed to choose materials to read according to their own interests and language proficiency level" (p. 14). However, there was no description in their article regarding how the learners' reading proficiency level was determined or even who determined their language proficiency level. There was no mention of percentages of unknown words or target percentages of known words for the learners. If it was merely left up to the students to select reading materials themselves at various levels and begin reading and determine for themselves whether or not the selected book is too difficult for them, then this process would be inefficient and most likely there would be students reading outside of their optimal reading level. In other words, while the students may be reading books they are interested in, they may also be selecting books that are far too easy and below their actual reading level. In addition, this process would most likely waste time that could otherwise be devoted to reading or another learning activity. According to the Extensive Reading Foundation (ERF) (2009a), research indicates that reading is at an 'instructional' level when the students know between 90 percent and 98 percent of the words on a page.

The third concern with Lee and Hsu's (2009) study is that they required their participants to write a paragraph or summary of their reading books, along with keeping a reading log. While their stated intention was to lessen the stress or burden placed on the learners by not requiring them to have as much accountability, I believe the amount of requirements placed on participants by Lee and Hsu in order to demonstrate whether or not they had read their books is still too high for many EFL learners. While I agree that it is important for the teacher to know who is on task or not, there are perhaps better methods available for determining this. Paragraph or summary writing may distract learners from the intended purpose of the reading, which is to benefit the learners' abilities. It is also in disagreement with Day and Bamford's (1998) suggestions for the top 10 principles of successful ER programs (see Table 1), specifically principles 4 and 5, which state that reading "is usually related to pleasure, information, and general understanding and … is its own reward." (p.8)

Table 1. *Day and Bamford's (1998, 2002) top 10 principles of successful ER programs*
Student read as much as possible
A variety of materials on a wide range of topics are available
Students select what they want to read
The purposes of reading is usually related to pleasure, information, and general understanding
Reading is its own reward
Reading materials are well within the linguistic competence of the students
Reading is individual and silent
Reading speed is usually faster rather than slower
Teachers orient students to the goals of the program
The teacher is a role model of a reader for students

The current study is based on the hypothesis that if EFL learners are exposed to the five conditions stated below, the learners will demonstrate significantly higher levels of improvement on measurements of writing than learners who are not exposed to the ER treatment.

1) A large variety of reading materials are provided.
2) Learners are placed in their optimal reading levels.
3) Learners experience a longer, more balanced duration of ER.
4) The duration of the daily ER is minimized to 15-20 minutes.
5) Less accountability is placed on the learners.

Method

Setting

The study took place in Taipei, Taiwan, at a middle-ranked private university. This university is unique in Taiwan in that it requires all of its non-English majors to participate in four years of EFL courses, when the standard at other Taiwanese universities is only two years for non-English majors. The university's EFL program is divided into eight levels—one for each semester. There is a standardized curriculum and course books that have been designed by the university's EFL teachers over the past 20 years. Students are placed into classes with classmates from their major, and there are no specific placement exams upon entering the university. Therefore, all of the classes within this program have mixed-ability students, ranging from extremely basic skills with almost no communication abilities to students with more advanced language skills.

Specifically, the study took place during the fifth and sixth level of the EFL program, or the entire third academic year. Each class met for two 50-minute back-to-back periods once per week. Each had the same researcher/instructor and received the exact same course curriculum with the only exception being the treatment of ER. Instead of the treatment, the control group (CG) received additional time to complete in-class activities (e.g., pair work or cooperative learning activities).

None of the classes adhered to the EFL curriculum established by the department, although both classes used the required vocabulary in both in-class activities and homework assignments. The

class curriculum was instead created by the researcher/instructor and specifically designed to follow a student-centered approach that focuses on authentic learning and the students' needs, abilities, interests, and learning styles instead of those of others involved in the educational process, such as textbook authors and administrators (Mermelstein, 2010). All of the classes can best be described as high-level communicative language classrooms with rich input focusing on all four language skills. All classes emphasized interaction as both the means and primary goal of the classroom. Thus, classroom activities took on several forms of pair and group work that required both negotiation and cooperation and were fluency-based in order to encourage the students to develop their self-confidence and authentic language skills. However, writing was also one of the main language skill requirements in all of the courses.

All of the students were specifically instructed in and assessed on paragraph writing and pre-writing activities (e.g., clustering). Approximately 50 minutes every other week was devoted to the development of the students' writing skills. This time was spent on organization and topic development, and also included the skill of paraphrasing. Several approaches were used, including direct instruction, pair work, group work, class demonstrations, using grading rubrics, and both peer and teacher assessments. However, at no time during the course of the academic year was there any direct grammar or vocabulary instruction.

Participants

Since the population of interest was undergraduate EFL learners in Taiwan, four 3rd year university classes of EFL learners were involved in the study. The total number of participants involved in the study was 211, with 60 male students and 151 female students. The participants can be defined as *convenience samples*, as they were already divided into four separate mixed-ability EFL classes by the university with the researcher randomly assigned as their instructor. The four classes were randomly designated as either part of the CG or the treatment group (TG) by the instructor/researcher prior to the beginning of the new school year, and the participants had had no previous contact with the teacher/researcher prior to the beginning of the school year. Due to university policy, the same instructor is not allowed to teach more than one class per academic level within the same department. For this reason, different class majors had to be selected for the study among the classes that had been assigned to the instructor/researcher. The four class majors participating in this study were accounting, information management, journalism, and statistics. The two former majors were randomly assigned as the CT and the two latter majors were randomly assigned as the TG.

The CG consisted of 33 male students and 71 female students. The TG consisted of 27 male students and 80 female students. All of the participants involved had previously studied EFL full-time for 8 years, including three years in both junior high school and senior high school and two prior years in the university's EFL program.

Reading Materials

The reading materials provided for this study were two separate graded reader series, the *Oxford Bookworms* and the *Penguin Readers*. For this study, only levels 1-6 were provided for the participants, as no starter levels were available. In total, there were approximately 600 graded

reader books available for this study, with approximately 100 books available for participants to select from at each level.

The graded reader levels are determined by the number of head words used by the publishers for each series. A headword is similar to a dictionary entry where a group of words share the same basic meaning (e.g., helps, helping, helpful, helpless). For this study, the graded reader scale provided by the ERF (2009b) was used. One of the main purposes for the ERF in creating the scale was to "have a uniform scale so that different series from different publishers could be compared in terms of difficulty level and placed together, and to provide a sense of uniformity between institutions" (ERF, 2009b). Therefore, students moving between one or two different series or publishers will remain aware of their actual reading level. The two graded reader series were selected due to their similarity in head words at each level.

Description of the Intervention

The intervention of this study is based on the intervention used in Mermelstein's (2013) study, which demonstrated a significant increase in the reading levels of Taiwanese university ESL students using ER. Therefore, the intervention of this study was an ER activity that took place in the classroom once per week as a sustained silent reading (SSR) activity. The duration of the SSR activity was on average between 15 and 20 minutes. The SSR activity was selected for the treatment for several key reasons. First, it is the belief of this researcher that an SSR activity can provide a more direct and personal interaction between the text and the individual learners. Second, it is a learner-centered activity that focuses on the needs and abilities of the individual learners. Third, SSR is supported by Day and Bamford's (1998, 2002) recommendation that *reading should be individual and silent*. And finally, this researcher believes that SSR is the only viable method of individualized reading that can take place in large mixed-ability classrooms.

In addition to the classroom treatment, the students in the TG were expected to continue doing extensive reading in their free time, with a minimum expectation of three pages being read from their graded reader books each day. Naturally, the participants were given permission, and even encouraged, to read more than three pages per day on their own. In addition to reading on their own, the TG participants were to keep track of their daily reading times and pages read on a specialized student reading record sheet. In order to maintain a balance among the CG and the TG of time spent outside of class devoted to English learning, approximately one hour of English homework each week was given to the CG. Homework for the CG consisted of cloze activities and intensive reading activities).

The overall framework of the intervention was also based upon all of Day and Bamford's (1998, 2002) top 10 principles (see Table 1) for conducting a successful extensive reading program. This study strictly adhered to all 10 principles. During week 1, a class discussion took place where the teacher/researcher explained the general purposes of conducting research in the ESL classroom and the benefits it can provide for students, researchers, and other teachers. Participants also received an informed consent form written in Mandarin Chinese, read and understood the brief explanations of the research study and the participants' rights, and signed it to indicate their agreement to participate in the study. During week 2, the participants were given and completed a pre-study graded reader reading level test to be used as a reading level

placement test (see Mermelstein,2013, for a full description). There was also a class discussion on the importance of reading in general and the overall benefits of reading within the proper reading level. During week 3, the participants were placed in their reading levels, based on a 95 percent understanding of the text level vocabulary. Instructions were given on how to read the graded reader books using inference instead of a dictionary. In addition, the participants were also taken to the school library and shown where to locate the graded reader books for checkout and given a sufficient amount of time to preview books and make their book selections for checkout. During week 3, the participants were also given a paragraph writing pretest to measure their starting writing ability level (The pretest is discussed in the Data Collection section). During week 4, the first SSR activity was initiated for approximately 20 minutes, followed by the participants filling in their graded reader record sheets. Weeks 5 through 17 of the first semester were similar, with the exception of week 9 due to the university-wide mid-term exams. Week 18 of the semester was the university-wide final exam week and participants were not able to meet in class for the SSR activity as well.

Following the completion of this final exam week, all of the participants in the study had a 5-week winter vacation and none met for in-class SSR. However, all of the members of the TG were given instructions to continue their outside class reading, with a minimum expectation of reading three pages from their graded reader books each day. In addition, they were to continue filling out their student reading record sheet (discussed in the Data Collection section).

Upon the start of the second semester, the SSR treatment began again and continued from week 1 through week 17 for the TG, with the exception of week 9 being used for mid-term exams. During week 17, a writing posttest was also administered (the posttest is discussed in the Data Collection section).

Data Collection Instruments

Two formal and two informal instruments were used to gather data, and the results were computed statistically. The two formal instruments were student writing samples, written without any feedback or revision, which served as pre- and posttests. Both writing samples were given as paragraph writing assignments. The pretest was given as an assignment during week three, prior to the start of the ER treatment, and the posttest was given during week 17 of the second semester at the end of the study. In both cases, the students were asked to do descriptive writing, with "Your Past Summer Vacation" and "Your Future Summer Vacation" as the topics for the pre- and posttests. Two raters were used to read and evaluate all of the writing samples for both the pre- and posttests. Both raters were native English speaking senior teachers who have been teaching Taiwanese ESL university students for more than 10 years each, one with an M.A. in TESOL, the other with a Ph.D. in TESOL. The paragraph writings were evaluated based on the criteria established by Jacobs et al. (1981), with five subscales: content, organization, vocabulary, language use, and mechanics. In addition, similar to Lee and Hsu's (2009) study, a sixth subscale was established using the total number of words written as a measurement of fluency. Following Jacobs et al.'s design, the total number of points assigned to each participant was determined by calculating the total number of points given for each component, with different weightings being assigned for each subscale. More value was placed on content, with 13-30 points possible. Next, both organization and vocabulary were worth 7-20 points, language use was worth 5-25 points,

and mechanics was worth 2-5 points.

The first of the informal data collection methods used was classroom observations, which Mason (1996) defined as "a method of generating data which involves the researcher immersing in a research setting, and systematically observing dimensions of that setting, interactions, relationships, actions, events, and so on within it" (p. 60). In other words, classroom observation may be a useful way to gather information on events that take place in language classrooms because it allows the researcher to observe specific behaviors at close range in order to understand the many variables of the classroom (Mackey & Gass, 2005). Therefore, the purpose of conducting the classroom observations of each reading class was to obtain valuable information regarding the implementation of the ER program and to assess student behavior (i.e., who was on task reading and who was not). For each SSR activity, the teacher maintained a record sheet noting which students were on task reading and which students were not. Specifically, this informal method of data collection was meant to lessen stress and learner accountability and to offset the workload of the students. Therefore, the students were not required to do additional work just to prove they were on task reading. As recommended by Day and Bamford (2002), the teacher was a role model for the students by reading during the SSR activity, or at least gave the appearance of reading throughout the SSR activity. In addition to reading, the teacher was actually observing students and inconspicuously placing marks on a seating chart indicating the student behavior as previously mentioned. Further, this entire manner of record keeping was enhanced by the intentional design of the student seating in the classroom, making it easier and less obtrusive for the teacher to gather data.

The second informal instrument used to collect data was the participants' graded reader record sheet, where daily information regarding their reading was logged. It included the title of the book they were reading, the dates that the participant actually read from their book, the number of pages read each day, and the amount of time it took them to read those pages. The purpose of the graded reader record sheet was threefold: (a) so that the participants could better track their own improvement and increase their internal motivation, (b) so that the teacher/researcher could track each participant's actual reading habits and identify participants who were not reading regularly, and (c) so that the teacher/researcher could track the participants' ongoing performance.

Results

Time on Task

Teacher observations of the treatment group took place in class during the SSR activity every week of the study. The results are represented in Table 2 as a mean percentage of time on task reading during the in-class SSR activity. On task means that the participants were fully engaged in reading their graded readers during the 15-20 minute SSR in-class activity, as measured by eye movement, pages turned, body language or movement, and the general facial expressions of the participants. The mean percentage of time on task for the treatment group was 94 percent. Off task time was generally due to the participants forgetting to bring their reading books to class rather than disruptive activities. Students who forgot their reading books were given more time

to complete other class activities not related to the study.

Table 2. *Mean scores for in-class teacher observations*

	# of students	Mean % of time on task	SD
Observation	107	94.13	9.79

Student Reading Record Sheets

The treatment groups' reading record sheets, as previously described, were informally reviewed every week of class during the study. Each week the teacher was able to review approximately half of the participants' reading behavior, as self-reported by the students on their reading record sheets. Students who demonstrated poor reading behavior were noted by the instructor and then followed up by an informal interview to discuss issues, assess the behavior, and counsel the participants when necessary. At the end of the study, the student reading record sheets were collected by the researcher; their self-reported reading data was analyzed and is represented in Table 3 as a percentage of the time on task reading outside of classroom (a percentage of the total days possible where the participants were reading outside of the classroom on their own time). Of the 107 participants in the treatment group who kept self-reported records of their reading, 89 sets of student reading record sheets were retrieved. The total mean percentage time of the participants' reading both inside and outside of the classroom, as self-reported, was 63 percent. This number represents 4.41 days per week of on task reading, meaning that the students reported reading at least 3 pages each day they read outside of class. Twenty participants (23 percent of the treatment group) self-reported reading every day. Due to the large number of students participating in this study, the total number of days involved in this study, the variability of reading speed among the participants, and the variability among the reading levels of the participants, the daily means of reading times and page amounts read for each participant are not included in this study.

Table 3. *Mean scores for self-reported student reading record sheets*

	# of students	Mean % of time on task	SD
Outside reading	89	63.21	27.55

A correlation analysis was conducted using Pearson's R to measure the correlation of the time spent outside of class reading the graded reader books (as self-reported by the participants) with the amount of time on task reading during the SSR activity. The results indicated a high, almost perfect, direct relationship of (0.90) to the amount of in-class SSR reading observed by the researcher with a *p*-value of 0.000 indicating a high significance level.

Writing Measurements

The two formal instruments were student writing samples, written without any feedback or revision, which served as pre- and posttests. Both writing samples were given as descriptive paragraph writing assignments, with "Your Past Summer Vacation" and "Your Future Summer Vacation" as the topics for the pre- and posttests (see Appendix for the complete writing prompts). The two writing topics were selected because they were as similar as possible without using the exact same topic.

At the beginning of the study, the results from the pretests of the CG and TG were analyzed using paired samples *t* tests. The results can be seen in Table 4. The results indicate there was no statistically significant difference in five out of the six categories. The only statistically significant difference was in the category of fluency, with p < 0.005. However, as with Lee and Hsu's (2009) study, the measurement of fluency was determined by the total number of words written by each participant. The numbers indicated by the CG and TG represent a mean of less than one word difference in the total number of words written by the participants and a standard deviation difference of less than one word as well. Therefore, while the numbers were statistically significant, they were not practically meaningful. In this researcher's opinion, the pretest results indicate two extremely comparable groups of participants in all six categories.

Table 4. *A comparison of pretest scores of the TG and CG for all subscales using mean scores*

	CG pretest	SD	TG pretest	SD	Sig.
Organization	13.519	3.687	14.039	3.547	0.266
Content	21.077	4.87	21.606	4.77	0.388
Vocabulary	12.702	3.813	13.48	3.813	0.117
Language use	14.202	5.662	15.26	5.273	0.144
Spelling/mechanics	3.182	0.845	3.356	0.7493	0.101
Fluency	52.058	11.004	52.99	11.427	0.003

At the end of the study, the results from the pre- and posttest were analyzed using a paired sampled *t* test. The first analysis was the CG's means scores for the pre- and posttests, and the results can be seen in Table 5. The next analysis was the TG's pre- and posttests, and the results can be seen in Table 6. The final analysis was the means of the CG's and TG's posttest scores, and the results can be seen in Table 7.

Table 5. *Control group's scores across all categories for pre- and posttests*

	Pretest	SD	Posttest	SD	Sig.
Organization	13.52	3.69	14.9	3.46	0.000
Content	21.08	4.87	21.79	4.52	0.000
Vocabulary	12.7	3.81	13.24	3.55	0.000
Language use	14.2	5.66	14.75	5.59	0.000
Spelling/mechanics	3.18	0.84	3.27	0.83	0.000
Fluency	52.06	11	54.73	8.51	0.040

Table 6. *Treatment group's scores across all categories for pre- and posttests*

	Pretest	SD	Posttest	SD	Sig.
Organization	14	3.51	15.54	3.25	0.000
Content	21.56	4.71	22.63	4.41	0.000
Vocabulary	13.42	3.78	14.71	3.14	0.000
Language use	15.15	5.24	17	4.54	0.000
Spelling/mechanics	3.32	0.76	3.82	0.72	0.000
Fluency	52.73	11.37	63.27	9.11	0.000

Table 7. *A comparison of posttest scores of the TG and CG for all subscales using mean scores*

	CG posttest	SD	TG posttest	SD	Sig.
Organization	14.903	3.457	15.577	3.288	0.116
Content	21.789	4.523	22.682	4.464	0.121
Vocabulary	13.24	3.548	14.789	3.146	0.001
Language use	14.75	5.586	17.163	4.498	0.001
Spelling/mechanics	3.269	0.827	3.846	0.721	0.000
Fluency	54.73	8.511	63.548	9.08	0.000

The results indicate that both the TG and the CG achieved significant gains on all six subscales with the TG making higher gains than the CG in all six categories. Using another multivariate analysis, the gains of the TG and the CG were then compared. The results indicate that the TG significantly outperformed the CG on five of the six subscales. The one subscale where the TG did not achieve significant gains over the CG was organization. This appears to be the effect of the fact that a majority of the in-class instruction regarding writing focused on pre-writing and other organization methods. Thus, the overall results of these analyses indicate support for the research hypothesis.

The inter-rater reliability for the pretest scores was 0.914, and the inter-rater reliability of the post-essay was 0.882. Therefore, both score ratings represent a high level of reliability.

Discussion

The purpose of this study was to determine if an enhanced design of an ER program would result in significant improvements in learners' writing skills. Therefore, this study first placed learners into reading levels at a 95 percent rate of understanding vocabulary, provided more access to reading books at the learners' levels where approximately 600 books were available to the participants to select from, increased the duration of the treatment period to 29 weeks, which was longer than in two of the three comparative studies, and decreased the amount of requirements on the learners by not requiring them to report or write about their reading books. The only difference in the two participant groups was the ER treatment. All of the participants were provided with the same in-class instruction and writing assignments and all of the participants followed the same course curriculum.

As the results indicate, the ER treatment appears to have had a large impact on the learners' performance, as the TG demonstrated significant improvement across all five subscales of the Jacobs et al. (1981) writing assessment and the sixth subscale added from Lee and Hsu (2009). The results also indicate significant levels of gains in four of the six areas being measured over the control group which did not receive the ER treatment. A comparison to five other ER and writing studies reporting the gains of their TGs over their CGs using statistical significance can be seen in Table 8. Only Lee and Hsu posted significant results in all of the subscales, but all of the studies posted significant results in at least two subscales. However, since Lee and Hsu and Tsang (1996) both used the same writing assessment as the current study, and since Lee and Hsu reported a comparison of Tsang's study using effect size, then a comparison with these two studies using the effect sizes reported in Lee and Hsu can also be made.

Table 8. *Comparison to five similar studies using statistical significance: TG gains over CG*

	This study	Lee & Hsu (2009)	Tsang (1996)	Hafiz & Tudor (1990)	Hafiz & Tudor (1989)	Lai (1993)
Organization	ns	sig	ns			
Content	ns	sig	sig			sig
Vocabulary	sig	sig	ns	ns	ns	
Language use	sig	sig	sig	sig	sig	
Spelling/mechanics	sig	sig	ns		sig	sig
Fluency	sig	sig		sig	sig	sig

Note. sig = readers significantly better than comparisons; ns = readers not significantly better than comparisons

A comparison of the three studies using effect sizes can be made in all of the subscales, except fluency, since Tsang (1996) did not use fluency as a subscale; it was created in Lee and Hsu's (2009) study. The effect size data of the three studies can be seen in Table 9. The current study established a small effect in organization and content, a medium effect in vocabulary and language use, and a large effect in spelling/mechanics and fluency. This data seems logical when comparing the gains of the CG and TG of this study, since the CG also made significant gains in all six subscales. Further, the two subscales where a smaller effect was observed are the two main components of the writing process that were instructed and practiced throughout the duration of the year in all of the courses. As stated earlier, all of the courses spent a considerable amount of time working on pre-writing activities and developing the main parts of a paragraph. It also seems logical that the other four subscales would receive a higher effect, since they can perhaps more easily be connected to the reading process (i.e., learning more vocabulary, learning how to spell more words correctly, and learning how to use these words more accurately in a sentence). With regard to the subscale with the highest effect—fluency—it could be attributed to the fact that all of the participants in both groups were actively writing throughout the school year and gained more confidence in their abilities. However, it is interesting to notice that the TG posted approximately 400 percent higher gains than the CG in this subscale.

Table 9. *Comparison to Tsang (1996) and Lee and Hsu (2009) using effect sizes*

	This study	Lee & Hsu (2009)	Tsang (1996)
Organization	0.19	1.32	0.32
Content	0.16	0.96	0.75
Vocabulary	0.22	1.15	0.32
Language use	0.44	1.15	0.63
Spelling/mechanics	0.73	0.75	0.14
Fluency	0.97	1.11	

When comparing the current study with Lee and Hsu's (2009), it appears that their study created a higher effect size among their participants. However, this was most likely due to the fact that the participants in their CG actually received lower scores on their posttests in four out of the six subscales. In other words, after one full academic year, their CG was actually performing worse than they were when they entered the course. Lee and Hsu should be applauded for their efforts, as their enhancement to the standardized course seems to have produced significant results with a large effect size.

Tsang (1996) also reported a small to large effect size across the subscales. However, similar to Lee and Hsu (2009), Tsang also reported that the CG of the study did not post significant gains on their posttest. In fact, they also posted significantly lower gains on one of the subscales—language use.

One important aspect of the current study, which no doubt had a positive effect on the outcome, was the learners' reading outside of class on their own time. Class time is often a factor of whether or not teachers feel they can add ER into their program (Takase, 2002). A direct comparison of the duration of in-class reading time with five of the previously mentioned studies can be seen in Table 10. The current study demonstrates that in order to produce significant learning results, only approximately 15 minutes of ER needs to take place in the classroom if it is supported by outside reading.

Table 10. *Comparison of total duration, total time spent reading*

Study	Duration (weeks)	Total time (hours)
Present study	29	7.25*
Lee & Hsu (2009)	30	25
Hafiz & Tudor (1989)	12	42
Hafiz & Tudor (1990)	23	90
Lai (1993)	4	50
Tsang	24	

Note. * = Total time of in-class reading

Since the other studies did not report on the time spent outside of the classroom reading, a direct comparison cannot be made with the overall time the studies' participants spent reading. However, since the learners of this study kept reading logs of both in-class and out-of-class reading times, it is possible to report that the overall mean time spent reading both in and out of class throughout the duration of the study was approximately 34 hours. This, then, is still less time than the majority of the other studies and supports the original hypothesis of this study, stating that if the duration of the daily ER is minimized to 15-20 minutes, and if the learners experience a longer, more balanced, duration of ER over the course of an entire school year, then they will demonstrate significantly higher levels of improvement on measurements of writing than learners who are not exposed to the ER treatment.

A much more general comparison can also be made between this study and other ER studies in Asia, thanks to Krashen (2007). He did a meta-analysis of 19 previously reported ER studies that had been published in professional journals or conference proceedings. Of the 19 studies, 12 of them were situated in Asia, and 10 of these were situated in Taiwan, like the current study. All 10 of the studies in Taiwan reported positive results in overall language proficiency, which led Krashen to state that the most obvious finding of the studies was that ER is consistently effective. However, Krashen also noted five factors which he believed played a role in the findings. These include the duration of the ER program, the length of time and frequency of the reading sessions, the extent of comprehension checking, whether or not the reading activity was encouraged, and whether or not the learners are under academic pressure. Therefore, without any previous intention, it appears that the current study lends support to Krashen's findings and also

supplements them with an indication that writing can improve with the addition of an ER program.

The current study used a one-year duration with regular weekly, if not daily, reading. The extent of measuring reading comprehension was done prior to the study with the pretest reading level placement and throughout the study through daily individual student assessment by the learners as they read their books. The reading activity was also encouraged every week in class when the teacher reviewed and discussed the individuals' reading record sheets with the participants. In addition, the learners were not under any heavy academic pressures to do any additional tasks related to the reading and were not expected to do any ER during the university exam weeks.

Conclusions

For many EFL teachers, the issue of promoting effective and proficient writing is important. Second language writing is a complex skill and teachers may need to use a variety of methodologies to best ensure their students' abilities improve over time. As this study demonstrates, the use of direct instruction and a wide variety of writing tasks and practice over the course of one academic year did result in significant improvement among the members of the control group. However, this study also demonstrated that the addition of ER provided greater results within the same time period, among a similar group of learners with similar cultural, language, and education backgrounds. Although the convenience groups contained four separate majors of study, it is a common belief among the EFL teachers at this university that students within these four majors are generally considered to be equal, in terms of English abilities, compared with students in other majors such as mass communication or law who tend to interact with English more and have stronger L2 language abilities. This researcher maintains the same belief.

Realizing the limitations of the classroom and the time available for teachers to directly interact with each student, ER may be able to help second language learners become more autonomous learners, especially in EFL environments where exposure to the target language may be limited. If an ER method is as effective as the results of this study and other studies suggest, then the implications for using an ER method as a secondary method of improving learners' writing abilities in the teaching of EFL may be extraordinary. Therefore, it seems logical for EFL teachers to seriously consider using the ER method to assist learners in their classrooms.

As one of the goals of this study was to try to establish an enhanced ER methodological design that can better serve EFL learners towards improvement, perhaps more researchers should be involved in this process. There is not a *one size fits all* method, and perhaps there will never be. However, the more research that is conducted in this thread, the more teachers will be able to streamline their lessons to better serve the learners and to use time more efficiently.

Limitations

Perhaps one of the limitations of this study was control over the time spent out of class reading or working on English homework assignments. Although the TG kept daily reading records of

the time they spent on task reading outside of class, the CG did not keep any records of time spent working on English homework. While the homework was designed and intended to keep the CG students engaged with English for approximately one hour each week, there is no way to know for sure the precise amount of time each student was engaged. Further, many of the TG students reported reading for longer than the minimum required amount of time. Additionally, the TG self reported reading for approximately 4.41 days per week, spreading out the time engaged reading among several days each week, as the study's hypothesis suggests. However, it is more than likely that the CG engaged in the English homework assignment from start to finish, only engaging with the assignment once per week. Therefore, even if the total amount of time spent each week engaged with English was the same among the two groups, the daily duration of time spent engaged with English was most likely not the same. However, based on the students' reading record sheet data, the study does indicate that ER is an effective, and most likely enjoyable, alternative use of time spent both inside and outside the classroom.

A second limitation was perhaps in the manner in which the pre- and post-writing samples were scored. In the design of the current study, the pre- and post-writing tests were scored at two separate times, one at the beginning of the school year and one at the end. It has been suggested that a blind rating system, where all of the writing samples are scored after the study has been concluded without the raters knowing which samples were pretests and which samples were posttest, would perhaps be a better design. This may be something future researchers should consider.

References

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

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language*, *26*, 29–48.

Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom.* Cambridge: Cambridge University Press.

Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, *14*, 136–141.

Elley, W. (1991). Acquiring literacy in a second language: The effects of book-based programs. *Language Learning*, *41*, 375–411. DOI: 10.1111/j.1467-1770.1991.tb00611.x

Elley, W., & Mangubhai, F. (1983). The impact of reading on second language learning. *Reading Research Quarterly*, *19*, 53–57.

Extensive Reading Foundation (2009a). Retrieved from http://erfoundation.org/wordpress/wp-content/uploads/2013/08/ERF_Guide.pdf

Extensive Reading Foundation (2009b). Retrieved from http://erfoundation.org/wordpress/graded-readers/erf-graded-reader-scale/

Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal*, *43*, 4–13. DOI: 10.1093/elt/43.1.4

Hafiz, F. M., & Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System*, *18*, 31–42. ERIC Number: EJ405466

Helgesen, M. (2005). Extensive reading reports: Different intelligences, different levels of

Processing. *Asian EFL Journal*, *7*, 25–33.

Huang, I. C. (1996). The relationship of free writing proficiency in second language acquisition. *Proceedings of the 13ᵗʰ Conference of the Republic of China TEFL*, Taipei: Crane, 133–143.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfield, V. F., & Hughey, J. B. (1981). *Testing ESL composision: A practical approach*. Rowly, MA: Newbury House.

Janopoulos, M. (1986). The relationship of pleasure reading and second language writing proficiency. *TESOL Quarterly*, *20*, 763–768. DOI: 10.2307/3586526

Krashen, S. (2007). Extensive reading in English as foreign language by adolescents and young adults: A meta-analysis. *International Journal of Foreign Language Teaching*, *3*(2), 23–29.

Lai, F. K. (1993). The effect of a summer reading course on reading and writing skills. *System*, *21*, 87–100. DOI: 10.1016/0346-251X(93)90009-6

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

Lee, S. Y. (2005). Facilitating and inhibiting factors of EFL writing: A model testing with SEM. *Language Learning*, *55*, 335–374. DOI: 10.1111/j.0023-8333.2005.00306.x

Lee, S. Y., & Hsu, Y. (2009). Determining the crucial characteristics of extensive reading programs: The impact of extensive reading of EFL writing. *The International Journal of Foreign Language Teaching*, *Summer*, 12–20.

Lee, S. Y., & Krashen, S. (1996). Free voluntary reading and writing competence in Taiwanese high school students. *Perceptual & Motor Skills*, *83*, 687–690. doi: 10.2466/pms.1996.83.2.687

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*, 532–543.

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, *25*, 91–102. DOI: 10.1016/S0346-251X(96)00063-2

Mason, J. (1996). *Qualitative researching*. London: Sage Publications.

Mermelstein, A. D. (2010). Matching teaching styles to promote student success. *Proceedings of the 2010 International Conference and Workshop on TEFL & Applied Linguistics* (pp. 351–360). Taipei: Crane Publishing.

Mermelstein, A. D. (2013). *The effects of extensive reading on 4ᵗʰ year Taiwanese students' reading level, reading speed, and perceptions of reading* (Unpublished doctoral dissertation). Tamkang University, Taipei, Taiwan.

Susser, B., & Robb, T. (1990). EFL extensive reading instruction: Research and procedure. *JALT Journal, 12*, 161–185.

Takase, A. (2002). The effects of extensive reading on Japanese high school students. Paper presented at the 2002 JALT Conference, Tokyo, Japan.

Tsang, W. K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics*, *17*, 210–233. DOI: 10.1093/applin/17.2.210

Waring, R. (2006). Why extensive reading should be an indispensable part of all language programs. *The Language Teacher*, *30*, 44–47.

Yamashita, J. (2013). Effects of extensive reading on reading attitudes in a foreign language.

Reading in a Foreign Language, *25*, 248–264.

Appendix

Writing Prompts for the Pretest and Posttest

Writing Prompt for the Pretest
Please write one paragraph describing this past summer vacation. You may select to write something general about the entire vacation or you may select to write about something specific that you did during your vacation time.

Writing Prompt for the Posttest
Please write one paragraph describing what you intend to do during this coming summer vacation. You may select to write something general about the entire vacation or you may select to write about something specific that you plan to do during your vacation time.

About the Author

Professor Mermelstein is an American certified teacher holding a M.A. in Teaching and a Ph.D. in TESOL. He taught ESL at levels K-12 in the USA and has been teaching EFL for 16 years at the post-secondary level in Taiwan. Areas of research include student-centered teaching, teaching methodologies, and reading. E-mail: merman2@hotmail.com