Effectiveness of CALL Vocabulary and Silent Sustained Reading on TOEIC Scores

Scott Miles
(Daegu Haany University)


TOEIC continues to be an important test in Asia and TOEIC preparation classes are as popular as they have ever been. Little research has been done, however, to determine the most effective ways to help students prepare for this test. This study sought to analyze the impact of explicit TOEIC instruction when supplemented with Silent Sustained Reading (SSR) and an online vocabulary learning program. Four combinations of these practices were tested: (1) explicit instruction supplemented by online vocabulary instruction, 2) a combination of explicit instruction, SSR and online vocabulary practice, 3) explicit instruction supplemented by SSR, and 4) explicit instruction only. Total time for instruction and practice were kept constant among the four groups. Due to between-group differences on pre-test TOEIC scores, an ANCOVA analysis was used. On total TOEIC scores, all groups made significant and statistically equal gains. However, on an analysis of the reading portion of the TOEIC results, group 2 (combination of explicit instruction, SSR, and online vocabulary practice) outperformed group 4 (explicit practice only group). The author concludes that the combination of all three practices had a synergistic effect on the reading portion of the test and led to superior gains.

I. INTRODUCTION

TOEIC continues to be a popular test in a number of Asian countries, Korea and Japan in particular, and is likely to stay that way in the foreseeable future (IIBC, 2005; Pan, 2010). In Korea, many companies require a minimum TOEIC score in order to apply, and a number of universities have TOEIC as a graduation requirement. TOEIC has recently added speaking and writing tests, but at the time of this writing scores on the reading and listening test remain as the chief requirement for most companies and institutions, and TOEIC classes in universities and private institutes which focus on the
reading and listening portion of the test continue to draw students. Yet despite the enduring importance of the TOEIC, there has been little direction from research as to how to best prepare students for the reading and listening test.

Although there is certainly some variety in how individual teachers conduct TOEIC preparation courses, several scholars have identified common instruction styles. Pan (2010) notes that test preparation courses tend to be teacher-centered, with students repeatedly practicing test questions and receiving feedback on their answers. Explanations often involve discussion in the native language about discrete grammar points and vocabulary items. Falout (2004) likewise observed that such practices tend to dominate TOEIC study in the Asian context. A number of publishing companies, both local and international, provide TOEIC study guides which tend to follow a similar format of explicit instruction and practice.

However, it may be an unsupported assumption that this kind of explicit preparation is the most effective way to prepare for the TOEIC. Robb and Erkanbac (1999), in one of the few controlled studies to measure actual gains on TOEIC from explicit instruction, found surprisingly weak results. They also point out that even if gains on the TOEIC are made after students take a test prep course, it may be difficult to pinpoint what percentage of gains came from increased test familiarity alone, and what percentage came from actual improvements in reading and listening ability that the TOEIC is claimed to measure.

A rigid focus on explicit practice and instruction for the TOEIC can also have negative side effects. Several scholars (Haladyna, Nolen, & Haas, 1991; Hamp-Lyons, 1998; Miller, 2003; Noble & Smith, 1994) have warned that tests such as TOEIC, TOEFL, and IELTS have undue prominence, leading educators and students to exert time and resources on the learning of test strategies and discrete item memorization which do not extend well to overall communicative competency. Students may learn to improve their test-taking skills, but are left unprepared to use the language outside of the classroom. The addition of activities in the TOEIC preparation course which can lead to improved TOEIC scores and develop language skills beyond test taking is certainly desirable. Even if explicit instruction for the test is useful for the improvement of TOEIC scores, this does not mean that entire study time needs to be limited to such. After strategies are adopted, actual reading and listening skills still need further development, and this may require activities beyond practicing test questions and receiving explicit feedback on test performance.

TOEIC instructors are generally hesitant, however, to include activities which are not directly related to TOEIC improvement, regardless of what other language benefits these activities might provide. Students enroll in TOEIC classes to increase TOEIC scores, and instructors who cannot deliver this result may find their jobs in jeopardy. TOEIC
instructors need evidence that the addition of activities which are not directly related to TOEIC development will also result in improvements on TOEIC scores.

This study will investigate the effects of replacing a portion of class time previously given to explicit instruction with activities that may not only have a positive impact on TOEIC scores, but also develop overall language proficiency as well. The activities examined in this study are two practices which have strong support in the literature for developing overall language proficiency: silent sustained reading (SSR), and explicit vocabulary practice through an online flashcard program. As one of the long-term goals of English educators should be to prepare students to be able to use English in the global community, language students would be well served should the inclusion of SSR and explicit vocabulary practice lead to equivalent (or even superior) gains on TOEIC scores in comparison to a purely explicit instructional approach.

II. LITERATURE REVIEW

1. The Effects of Explicit TOEIC Instruction and Practice

To date, only a handful of studies have been conducted on the effects of TOEIC test preparation courses. Boldt and Ross (1998) analyzed existing data collected from a number of companies and training institutes in Japan on the impact of four specified factors (training objectives, instructor background, teaching materials, and class size) on TOEIC score gains. The review found that classes with a combination of video materials, trained instructors, and medium-sized classes had the best effect. However, though the results of these programs were measured on TOEIC scores, it appears that the content and objectives of the courses were focused on improvement in general business English, as the materials used in the classrooms were composed primarily of general 4-skill development English texts, current events/news articles, video, and business simulation, rather than materials designed specifically for TOEIC improvement.

As noted earlier, Robb and Ercanback (1999) conducted one of the few studies measuring the results from explicit TOEIC test preparation. Two samples of students (English majors and non-majors) at a Japanese university were divided into three groups: 1) TOEIC Preparation, 2) Business English and, 3) General (four-skills) English. The study found that the only significant gains on the post-test scores were found for the non-English majors on the reading portion of the test only, while the other groups made no progress during the study period (8 months). Indeed, several groups actually experienced a slight decrease on some scores. The authors concluded that TOEIC prep classes are of questionable benefit for freshmen university students, but they note that
this may be attributable to the general poor attitude of Japanese freshmen students towards general education subjects, rather than as an indictment of the instruction methods utilized in the study.

Falout (2005) likewise found that explicit study and practice of the TOEIC had somewhat disappointing results. Falout followed the progress of 35 university students in Japan in a TOEIC study group over a 4-year period and in his analysis concluded that students who study daily, actively, top-down, and diversely (participating both in group and individual study) made the best progress. Conversely, Falout identified several factors which were perceived as having little if any impact on improving TOEIC scores, including explicit study of the test itself, rote memorization of vocabulary, discrete point grammar study, and a bottom-up approach to listening (straining to catch every word and listening to the same sentence repeatedly), factors which are quite common in explicit TOEIC instructional settings. However, the study did not include statistical analyses, so it is unclear if these observations are valid.

Pendergast (2010) found far more positive results from explicit TOEIC instruction. He found that Japanese students majoring in English gained an average of 135 points on the TOEIC test (increase in mean average 265 to 400) over two years of instruction (estimated 500 hours) which combined traditional instruction with reading. As there were no comparison groups in the study, it is unclear as to what portions of the gains can be attributed to explicit instruction and reading, as well as outside factors such as independent study, additional study in private institutes, and so on.

A tentative conclusion from this handful of studies on TOEIC score development is that explicit study for TOEIC may have some effect on TOEIC scores, but the active study of English in general may produce similar, if not superior, results. For instructors and students of TOEIC, there is at this time little direction as to how to prepare for the test.

2. Vocabulary Knowledge and TOEIC

One perhaps unavoidable aspect of succeeding on TOEIC is to develop a strong knowledge of vocabulary. Nation and Meara (2002) claim ‘there is a relatively close relationship between how many words you know, as measured on the standard vocabulary tests, and how well you perform on reading tests, listening tests and other formal tests of your English ability’ (p. 50). Qian (1999) made a comparison of Vocabulary Levels Test scores and the TOEFL reading comprehension section and found the two tests had a strong correlation of .78. Similarly, Beglar and Hunt (1999) found a correlation of .71 on the comparison of 2000 Word Level Test scores and TOEFL scores of 496 EFL students in Japan.
On the TOEIC, Kanzaki (2010) found a strong overall correlation (.64) between vocabulary knowledge and TOEIC scores. The reading section had a high correlation of .76, while the listening section had a weaker correlation of .39. Kanzaki suggested that this is because the vocabulary tests used did not have an audio component, and further, the listening portion has less of a demand on vocabulary knowledge than the reading portion. Additionally, in further analyses Kanzaki discovered that test familiarity may have accounted for the strength of the correlation. When students with low test familiarity were analyzed, the correlation of vocabulary knowledge to overall test results dropped to .49. Mizumoto and Takeuchi (2008) found that vocabulary learning strategies had the greatest influence on TOEIC scores, showing that students with high scores tended to have ‘...clear goals and attended to vocabulary learning strategies in conscious, coordinated, and structured manners’ (p. 17).

Laufer (1989, 1992) and Nation (2001) have suggested that learners need a vocabulary coverage level of 95 percent in order to reliably read texts. Chujo and Genung (2004), in an analysis of TOEIC test vocabulary coverage, found that knowledge of roughly 3714 words was necessary to comprehend 95% of the average TOEIC test. Later, Chujo and Oghigian (2009) analyzed a corpus made up of a number of TOEIC tests and found that in order to gain 95% text coverage, students would need a vocabulary of at least 4000 words (3000 word families). Without a strong knowledge of these words, students may inevitably struggle to comprehend the majority of the passages on any given TOEIC test. Chujo and Nishigaki (2003) noted that the vocabulary taught in Japanese high school textbooks was insufficient in preparing students for TOEIC tests, and stressed the need for further vocabulary development.

The question remains, however, as to how instructors can assist students in developing this needed vocabulary knowledge. At this time, there is not much direct evidence that the explicit study of vocabulary common to TOEIC leads to substantial gains on TOEIC tests. Tanimura and Utiyama (2006) compared three groups of students: one group using only vocabulary word lists, another simply reading texts utilizing high frequency TOEIC vocabulary, and a third utilizing both methods. The group using only texts outscored the other groups, with the group using the combination of texts and explicit vocabulary study outperforming the group that only studied vocabulary. The results suggested that simply reading may be the best way to gain vocabulary knowledge for TOEIC improvement. A second research question in the study examined the gains on the TOEIC test after studying TOEIC words and found a very slight but statistically significant increase on the test. However, as there was no comparison group, it is unclear whether this modest gain came as a result of the vocabulary study or other factors such as general English improvement or increased test familiarity. As noted earlier, Kanzaki (2010) found different correlation strengths depending on the students’ familiarity with
the TOEIC test.

3. CALL for Vocabulary Learning

Recently there has been strong interest in using CALL programs for vocabulary learning. Students need to learn a great deal of vocabulary, yet most teachers are unable to devote enough time in class for these goals. CALL programs for vocabulary development outside of class, then, can have great appeal for teachers and learners. Not only can CALL vocabulary programs provide vocabulary exercises, but also systematic repetition of learned words in order to maintain retention. Studies in memory research and vocabulary learning (i.e., Bahrick, 1984; Miles & Kwon, 2008) have found students learning vocabulary through spaced repetitions have far superior long-term retention in comparison to students learning in massed repetition conditions (intensive study over a short period of time). Pimsleur (1967), proposed an expanding repetition schedule for effective retention (see Table 1) based on the principle that with each repetition, the learner can wait progressively longer periods of time before needing further review. However, practical application of these principles may be difficult, as it would not be easy for most students to remember when to review different sets of words. The advantage of a CALL vocabulary system, then, is that the computer can remember when studied words are due for the next review session.

TABLE 1
Pimsleur's Memory Schedule

<table>
<thead>
<tr>
<th>Study Session</th>
<th>Waiting time before next review session</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 seconds</td>
</tr>
<tr>
<td>2</td>
<td>25 seconds</td>
</tr>
<tr>
<td>3</td>
<td>2 minutes</td>
</tr>
<tr>
<td>4</td>
<td>10 minutes</td>
</tr>
<tr>
<td>5</td>
<td>1 hour</td>
</tr>
<tr>
<td>6</td>
<td>5 hours</td>
</tr>
<tr>
<td>7</td>
<td>1 day</td>
</tr>
<tr>
<td>8</td>
<td>5 days</td>
</tr>
<tr>
<td>9</td>
<td>25 days</td>
</tr>
<tr>
<td>10</td>
<td>4 months</td>
</tr>
<tr>
<td>11</td>
<td>2 years</td>
</tr>
</tbody>
</table>
There is some evidence that these types of online flashcard vocabulary systems can be effective for vocabulary development (Miles & Kwon, 2008), but as of yet no study has found a direct link between online flashcard vocabulary learning and improvements on the TOEIC. Phillips (2011) compared two groups in TOEIC preparation classes: one group using an online vocabulary program utilizing a repetition schedule, and the other group having only in-class instruction. Students using the online vocabulary program improved on the TOEIC by an average of 77 points. In comparison, the control group increased on the TOEIC score by 65 points. No statistics were provided by Phillips, however, to see if the differences were statistically significant. Considering the relative small difference in scores (12 points on a 990 point test) and the small sample size of students (27), it is unlikely that the differences were significant. However, the results do present the possibility that online vocabulary programs could have a positive influence on TOEIC scores.

Agawa, Black, and Herriman (2011) conducted a similar study on TOEIC scores with students (76) using an online vocabulary program for a period of eight weeks, with students learning an average of 262 words common to TOEIC tests during this period. The authors compared the progress made by these students to the progress made by students in previous years taking the same course but without the online vocabulary component. Though the students gained 55 points on the test, which was as high or higher than the progress made by students in previous courses, the overall mean differences were not significant. The authors concluded that students would need to pass a higher threshold of newly learned vocabulary in order to affect their TOEIC performance.

Thus, the few studies looking at direct vocabulary study through online CALL programs suggest that they may be beneficial, but as of yet there is no clear evidence of improved TOEIC performance in comparison with students who only participated in standard TOEIC classes.

4. Free Voluntary Reading: Silent Sustained Reading

In addition to online vocabulary study, another promising practice for TOEIC score improvement is free voluntary reading, namely extensive reading and silent sustained reading. Silent sustained reading (SSR) is the practice of allotting a certain period of time during class (typically 10-20 minutes) for student self-selected silent reading. In most programs, teachers provide a variety of interesting books at the students’ reading level (thus requiring no assistance to read) and the students simply read independently during the allotted time.

McCracken (1971) was one of the first to lay out the basic concepts of SSR.
• The whole class reads
• A regular segment of time in the class is allotted for SSR
• Students self-select reading materials
• Students read one book for the allotted time
• The teacher participates as well, modeling the activity
• No records are kept.

Pilgreen (1994) added the factors of ‘book appeal’ (providing access to books which are of interest to the students), ‘conducive environment’ (quiet, comfortable environment for reading), and ‘non-accountability’ (no burdensome reports, quizzes, etc.).

SSR has much in common with the practice of extensive reading. The primary difference between SSR and extensive reading is that an extensive reading program attempts to give the students far more exposure to texts. SSR is relegated to the classroom, while extensive reading can be done inside the class (in which case it is indistinguishable from SSR), outside the class as homework, or a combination of both.

There is strong evidence that fluency reading activities such as SSR and extensive reading lead to the development of overall reading skills as well as or better than the traditional approaches to reading which emphasize assigned texts, direct instruction of vocabulary and grammar, and comprehension tests (Manning, Lewis & Lewis, 2010; Wiesendanger & Birlem, 1984). Indeed, studies following student progress for over one year tend to show an advantage for the groups involved in free voluntary reading (Krashen, 2003).

The theory behind SSR is that reading fluency is developed through time spent actively engaged in the task of reading: we learn to read by reading. Extensive exposure to texts is necessary to build up sight word vocabulary and automaticity in parsing sentences. As Grabe (2002) claims, ‘given that reading efficiency is dependent on rapid and automatic word recognition and a large recognition vocabulary, extensive exposure to L2 texts through reading is the only learning option available to L2 students’ (p. 56). Intensive reading practices, in which the students read short and challenging texts followed by explanations by a teacher and intensive study of vocabulary and grammar, simply do not provide students with enough opportunities to get the needed massive exposure to texts. In typical TOEIC preparation courses, intensive reading practice is the standard (Falout, 2004; Pan, 2010).

Several studies have shown that SSR in the second language classroom can lead to substantial language gains. Lee (2006) showed impressive gains in general vocabulary development and language ability (as measured on a cloze test) for a group which engaged in SSR over a period of one year, in comparison to classes which followed the
Effectiveness of CALL Vocabulary and Silent Sustained Reading on ... 347

regular curriculum. Smith (2010) found that by adding an SSR component to a reading class, students made better reading progress than those who only had traditional reading instruction. 'By replacing one element of intensive instruction with sustained silent reading, students did not lose ground as some may fear, but in fact outgained their counterparts.' (Conclusion section, para. 1)

A similar study was conducted by Petrimoulx (1988) on ESL students at a university in the United States, with students in the experimental group engaging in SSR for 10 minutes a day for 15 weeks. In comparison to control groups with standard intensive reading instruction, SSR students had higher reading comprehension and vocabulary scores at the end of the study, but the differences were not statistically significant. Nonetheless, this study shows that class time 'lost' to SSR did not lead to any disadvantage for the students in comparison to students receiving standard reading instruction.

There have been few studies examining the effect of SSR and extensive reading practices on TOEIC scores. Storey, Williamson, and Gibson (2006) found an increase in TOEIC scores resulting from extensive reading. The group adding extensive reading (one graded reader a week) to their explicit TOEIC instruction course had larger gains than the comparison group which did pure explicit TOEIC study, but again the differences fell short of statistical significance.

SSR in class can be a strong challenge to teachers and learners who are accustomed to a teacher-centered classroom. The idea of giving valuable class time to an activity which could be done as homework can strike educators and students as unnecessary. However, the research to date gives compelling evidence that SSR can provide equal, if not better, results than teacher-centered reading instruction. As the TOEIC reading section demands rapid reading, SSR seems to be a promising addition to TOEIC preparation courses.

5. Summary of Literature Review

The literature to date shows little direction for TOEIC preparation courses. The effects of explicit instruction and practice are largely untested. Online vocabulary study and SSR are promising additions to TOEIC study, but as of yet there is little direct evidence that adding these practices to a TOEIC preparation course will have a strong impact on TOEIC scores.

It is important to keep in mind that there may be a problem with focusing on an 'either/or' approach to classroom practices. In many cases, a combination of methods or practices may provide the best results. This study, then, will not only seek to measure the effect of explicit instruction, online vocabulary practice and SSR on TOEIC
performance, but also explore several combinations of said practices to see which, if any, produces the best results on TOEIC improvement.

III. METHODOLOGY

To investigate the impact of explicit instruction, SSR and online vocabulary study on TOEIC scores, four intact TOEIC classes taught in the fall semester at a private university in South Korea were selected. The classes were relegated into four experimental groups (see Table 2).

1. Participants and Course Details

All participants in the study were first year students at a private university in South Korea taking a mandatory TOEIC course (‘TOEIC 2’) in the fall semester of 2011. All of the students had previously taken a TOEIC preparation course in the spring (‘TOEIC 1’). Thus, all students were generally familiar with the TOEIC test and basic test-taking strategies.

The TOEIC classes were divided by majors. The classes in this study include majors in nursing, cosmetic science, special education, and physical therapy. All groups met once per week for 2-hour classes in the late morning or early afternoon. The number of participants per Group ranged from 34-44. All students were Korean; the majority aged 19-20. The English proficiency levels of the students varied, but the majority could be considered high-beginner to low-intermediate. The majority of students in each group were female (see Table 2).

The courses were taught by three experienced instructors (referred to as instructors S, L, and K). All were female Korean instructors who had taught TOEIC courses for over 4 years. The courses were generally conducted in the L1 (Korean), a standard practice in TOEIC instruction classes in Korea. All groups followed the same textbook, Longman Preparation Series for the New TOEIC Test: Introductory Course (Lougheed, 2006) by Pearson-Longman publishers. Interviews were conducted with each teacher to find out how they taught their courses. All three teachers generally followed the same format common to most TOEIC instruction contexts (see Introduction). The only variation of note (aside from the experimental interventions) was that Instructor L, teacher of Groups 2 and 3, added a short (5 minute) dictation activity as a warm-up for listening practice.
TABLE 2
Study Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Female</th>
<th>Male</th>
<th>Major</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (+voc)</td>
<td>34</td>
<td>30</td>
<td>4</td>
<td>Nursing</td>
<td>S</td>
</tr>
<tr>
<td>Group 2 (+voc +SSR)</td>
<td>37</td>
<td>32</td>
<td>5</td>
<td>Nursing</td>
<td>L</td>
</tr>
<tr>
<td>Group 3 (+SSR)</td>
<td>44</td>
<td>36</td>
<td>8</td>
<td>Cosmetic science</td>
<td>L</td>
</tr>
<tr>
<td>Group 4 (control)</td>
<td>39</td>
<td>27</td>
<td>12</td>
<td>Physical therapy/Special education</td>
<td>K</td>
</tr>
</tbody>
</table>

2. Study Treatments

There were three types of practices examined in this study: explicit instruction and practice, SSR, and supplemental online vocabulary practice. All four groups received explicit instruction and practice. Group 1 (+voc) received online vocabulary practice to supplement the explicit instruction/practice. Group 2 (+voc, +SSR) supplemented explicit instruction and practice with online vocabulary and in-class sustained silent reading. Group 3 (+SSR) supplemented their explicit instruction/practice with sustained silent reading. Group 4 (Control) received only explicit instruction and practice. Total time for study was held constant for all four groups, with all groups spending roughly the same amount of time for homework.

3. Explicit Instruction

All groups had explicit TOEIC instruction as the main part of the course. Each class followed the same course book and the same schedule (all units in the book were covered by all the instructors). Lessons consisted primarily of taking short practice tests on a specific question type, followed by an explanation of the answers, and then a lecture of relevant language items and test-taking strategies. Explanations of all grammar points, test strategies, and answers to practice questions were given in Korean.

Group 1 (+voc) and Group 4 (Control) had two hours of explicit instruction and practice in each class. Group 2 (+voc, +SSR) and Group 3 (+SSR) had 1.5 hours of explicit instruction in each course, with the other 30 minutes given to SSR. Groups 3 (+SSR) and 4 (Control) had further explicit practice as homework. This was in the form of doing practice quizzes and exercises, and was estimated to take about 40 minutes each week.
4. Sustained Silent Reading (SSR)

Group 2 (+voc, +SSR) and Group 3 (+SSR) (both taught by Instructor L) replaced 30 minutes of explicit instruction in class with an SSR activity. This was done for the final 30 minutes of each class.

In Week 3 of the course, the students in Groups 2 and 3 were introduced to the practice of SSR. The instructor described the activity and briefly discussed the benefits of such reading practices. This was done to ensure that the students understood the purpose and utility of this kind of reading, and to understand how they should approach the readings and find the right level for them. The teacher brought in a collection of graded readers (see below), and allowed the students to select the reader they wished to read for that period. This was continued through Week 14, with one week missed for midterms and another week missed for school holiday. The students thus had a total of ten SSR sections, with most finishing 9-10 readers in total.

Footprint Readers were used in the study. The Footprint Reading Library series consists of nonfiction books produced by Thompson Heinle. The Footprint Reader series has eight levels, beginning at the 800 word level (vocabulary was mostly restricted to the 800 or most common words in the English language) and proceeding to the 3000 word level. The average number of words per book varies according to the levels of the book, but the range of word counts of the books used in this study was from 600-940 for main content, and another 300-335 words of supplementary text in the back of the book.

Footprint readers were selected for the following reasons:

- The length of each book is short enough for students to finish in one SSR session. This is not a requirement of SSR, but was desirable for this study as the logistics of ensuring that a student could get the same book that was read in the previous class would be difficult for the teachers to manage due to the large number of students in each class.
- The books include a glossary of the more difficult words, enabling students to read without having to pause to look up new words in a dictionary.
- The books have more graphics and illustrations than most other readers, and thus are likely to be more attractive and interesting to the students.

30 minutes were given for SSR in each class. Generally, students took 5 minutes to select their book for that day and to begin reading. Upon completion of the reading, students filled out a short form (see Appendix C) indicating how much they liked the story and what they thought of it. Students were encouraged to write in Korean in order to make writing the report less of a burden and keep this portion of the activity brief.
Most students finished the book (including the supplementary material at the end of the book) and the report at the 20-25 minute mark.

With SSR taking 30 minutes of each lesson, the instructor of Groups 2 and 3 was forced to reduce the amount of practice from the course book. This was done equally with listening and reading. The instructor still covered all of the units, but the amount of practice in each unit was reduced.

To keep classroom procedure logistically simple for the instructor, only two levels of Footprints were made available to the teacher: the high-beginner (800 word level) and the pre-intermediate (1000 word level). The majority of the students indicated that these levels were somewhat easy, making them appropriate for SSR. There were only a handful of students (3 of 81 students in both groups) who found the books consistently difficult. There were slightly more students (6 of the 81 students) who found the books consistently easy. To stay true to the principles of SSR, it would have been better to allow these students to have access to books at the appropriate levels, but the students were satisfied with the books and the instructor preferred not to further complicate class procedure by offering different level readers.

5. CALL Vocabulary Programs

Two groups in the study, Group 1 and Group 2, used an online vocabulary program (Praxised.com) for the homework component of the class. In each study session, students were presented with 10 new words, followed by a variety of exercises to further strengthen their knowledge and retention of the words. Words from previous sessions were recycled according to a memory schedule to ensure retention. Students were assigned to complete 40 sessions of the program by the end of the semester (roughly 3 15-minute sessions per week).

New words in each session were initially presented with 1) a definition of the word in English accompanied by a Korean translation of the definition, 2) a model sentence using the word in English, accompanied by a Korean translation of the sentence, and 3) a list of 4-5 common collocations and colligations of the word (with Korean translations).

After a word was introduced, it was followed by 2 review exercises of the word that came up later in the same session. Review exercises consisted primarily of flashcard activities, from L1-L2 and vice-versa. Other activities included hearing the word and matching the Korean translation, and completing partial cloze exercises cued by Korean translations. Further reviews of a given word appeared in later sessions. Specifically, words introduced in one session would be reviewed on the following session, then again 7 sessions later, and finally once more 17 sessions after the initial session. This
repetition schedule is based loosely on the Pimsleur’s (1967) model (see Table 1).

The words used in the program were taken from research conducted by Chujo (2003). Chujo analyzed a number of TOEIC tests and created a list of words which appeared frequently on the tests (excluding words from the most frequent 2000 words in English). From Chujo’s list, the researcher selected 400 high frequency words for the online program. Students were required to do a total of 40 sessions, which would provide the study of all 400 words. Students doing 30-45 sessions were retained in the study, meaning the average student had exposure to about 380 words. Students averaged 15 minutes per session on the site (standard deviation of 4:45 minutes). The total time spent on the site averaged 10 hours.

6. Test Administration

As noted in Section III-1, all students in the study had taken a TOEIC class (‘TOEIC 1’) in the previous semester. At the end of the semester, the students took a mock TOEIC test (listening and reading), which accounted or 10% of their grade. All classes take the same test, which is changed each semester.

The original plan of the study was to give a similar mock TOEIC test in the second week of the fall semester to serve as the pretest of the study. This test was conducted in September of 2011, but it became apparent that there were serious problems with the test results. The instructors of the classes all reported that a significant number of students were visibly distracted during the test and left many questions blank. This is likely due to the fact that the students knew that the results of this test would not have an effect on their grade. As all of these students had taken a similar mock TOEIC test at the end of the previous semester (June, 2011) for their TOEIC 1 course, the researcher was able to compare the means of the test taken in early September to those of the test taken in June. The results showed a steep average decrease on mean scores of 100-140 between the two tests. As it is unlikely that the students’ TOEIC ability dropped by such a large margin in the 2-month time period between these tests, these results confirmed the observations of the instructors that the students were not fully engaged in the test given in September.

Thus, the researcher made the decision to use the test scores from the end of the spring semester (June, 2011) as the pre-test measure for the study, as there is more confidence that this test reflected the students’ actual TOEIC ability than the test given in September. However, as there was roughly a 2-month period between the test scores recorded in June and the beginning of this study, there is the risk that during that period some students may have received additional English study, and thus increased their actual TOEIC ability before the experimental methods could begin. To counter this
threat to the validity of the study, the students were surveyed to find out which students had indeed received additional English study over the summer break. Students who sought out additional instruction over the summer break were removed from the analyses of the study (see below for more information on the surveys).

Thus, despite this unusual procedure, the researcher had a reasonable degree of confidence that the scores used for the pretest measure could be used reliably for the purpose of the study. Certainly some students may have experienced a slight drop in performance over the 2-month break, but this would likely apply equally to all students in the study.

The post tests were taken in the final week of the course (early December, 2011). Just as the tests used for the pretests, the scores on the final TOEIC test made up 10% of the students’ grades, giving the students incentive to do their best on the test.

As evident from the pretest results (see Table 3), not all groups began at an equal level, with the two nursing major classes (Groups 1 and 2) far above the other two groups. These differences were confirmed by an ANOVA \[ F(3, 143) = 9.99, p < .01 \] performed on the total TOEIC scores. To account for the pre-test differences in mean scores, an ANCOVA analysis was used on the post-test results with the pre-test mean scores as the covariate. The independent variable was the intervention type (Groups 1-4), and the dependent variable was the post-test scores. The covariate for each analysis was checked to ensure there were no violations of the assumptions of normality, linearity, and the homogeneity of variances and regression slopes. All statistical analyses were conducted on SPSS version 17.0.

7. Survey

A survey (see Appendix A and B) was conducted around the midterm period. The purpose of the survey was to find out if students had received previously (over the summer break) or were receiving at that time any English instruction in addition to the TOEIC course, in college or out of college. All students who had taken English courses over the previous summer or who were taking additional classes during the study period (autumn of 2011) were excluded from the analysis of the study. The results showed very few students (six out of the all the participants) had taken any English classes over the summer, though a number of students (18) were currently taking English courses. These students were removed from the analyses of the study.
IV. STUDY RESULTS

Table 3 gives the descriptive statistics for the total pre and post-test TOEIC scores, including overall gains from pre- to post-test.

### TABLE 3
**Total TOEIC Score Means and Gains**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre-test M</th>
<th>SD</th>
<th>Post-test M</th>
<th>SD</th>
<th>Total Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (+Voc)</td>
<td>34</td>
<td>521.8</td>
<td>117.2</td>
<td>631.8</td>
<td>124.6</td>
<td>110</td>
</tr>
<tr>
<td>Group 2 (+Voc, +SSR)</td>
<td>37</td>
<td>495.5</td>
<td>118.8</td>
<td>646.8</td>
<td>123.9</td>
<td>151.3</td>
</tr>
<tr>
<td>Group 3 (+SSR)</td>
<td>37</td>
<td>396.6</td>
<td>98</td>
<td>518.2</td>
<td>131.6</td>
<td>121.6</td>
</tr>
<tr>
<td>Group 4 (Control)</td>
<td>39</td>
<td>418.3</td>
<td>120</td>
<td>539.4</td>
<td>141.4</td>
<td>121.1</td>
</tr>
</tbody>
</table>

Group 2 (+Voc, +SSR) made the most progress on TOEIC scores with average gains of over 151, while Group 1 (+Voc) made the least progress with an average 110 gain on the post-test. However, after adjusting for the pre-test scores, the ANCOVA analysis found no significant differences between the groups \([F(3,142)=1.97, p=.12, \text{partial eta-squared=.04}]\).

Each section of the test results (reading and listening) was also analyzed with an ANCOVA to see if the treatments had any significant effects on those individual skills. Table 4 shows the results of the listening portion of the tests.

### TABLE 4
**TOEIC Listening Score Means and Gains**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre-test M</th>
<th>SD</th>
<th>Post-test M</th>
<th>SD</th>
<th>Total Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (+Voc)</td>
<td>34</td>
<td>265.7</td>
<td>77</td>
<td>344</td>
<td>74.2</td>
<td>78.3</td>
</tr>
<tr>
<td>Group 2 (+Voc, +SSR)</td>
<td>37</td>
<td>261.9</td>
<td>79.5</td>
<td>355.5</td>
<td>78.2</td>
<td>93.6</td>
</tr>
<tr>
<td>Group 3 (+SSR)</td>
<td>37</td>
<td>221.6</td>
<td>65.8</td>
<td>297.6</td>
<td>82.9</td>
<td>76</td>
</tr>
<tr>
<td>Group 4 (Control)</td>
<td>39</td>
<td>221.8</td>
<td>67.8</td>
<td>304.2</td>
<td>88.3</td>
<td>82.4</td>
</tr>
</tbody>
</table>

On the listening portion of the test, Group 2 (+SSR, +Voc) again made the largest gains, but the ANCOVA failed to find statistically significant differences between the mean scores of the four groups \([F(3,142)=1.41, p=.24, \text{partial eta-squared=.03}]\).

Table 5 shows the results of the reading portion of the tests.
Once again, Group 2 (+Voc, +SSR) had the highest gains among the groups. An ANCOVA performed on the reading portion of the TOEIC showed a difference between the groups \[ F(3,142)=2.95, p=.03 \]. The partial eta-squared score was .06, indicating a moderate effect size. To determine where the differences lied among the groups, a post-hoc analysis (Bonferroni) was conducted. The differences of mean scores on the reading portion of the test for Group 2 (+SSR, +Voc) were statistically significant over Group 4 (Control). The difference in means between Group 2 (+SSR, +Voc) and Group 3 (+SSR) fell short of statistical significance \( p=.13 \).

### V. DISCUSSION AND CONCLUSION

Group 2 (+SSR, +Voc) had the highest gains on the overall TOEIC score, with superior gains on both the listening and reading portions of the test. However, according to the results of the ANCOVA, the only statistically significant results were found on the reading portion of this test, in which Group 2 outperformed Group 4 (Control). On all other measures, none of the differences among the groups reached statistical significance.

One interpretation of these results is that the combination of explicit study, time for SSR, and online vocabulary study had a synergistic effect resulting in superior TOEIC gains on the reading portion of the test. The sole addition of either SSR or online vocabulary support was not enough to make a significant difference compared to explicit TOEIC instruction and practice, at least within the relatively short duration of this study.

A comparison of Group 3 (+SSR) and Group 4 (explicit study only) largely coincides with the results of Storey, Williamson, and Gibson (2006). The addition of an extensive reading component (SSR) lead to higher gains in comparison with the group which only experienced explicit study and practice, but the results did not reach significance. Likewise, a comparison of Group 1 (+voc) and Group 4 (explicit study only) confirmed
the results of previous studies failing to find a significant advantage to adding CALL vocabulary practice to TOEIC instruction (Agawa, Black, & Herriman, 2011; Phillips, 2011; Tanimura & Utiyama, 2006). However, as all of these studies were relatively short term (one semester), the full potential of online vocabulary practice is still unexplored. For explicit online vocabulary learning to have a strong effect on TOEIC improvement, a certain threshold of newly added vocabulary may need to be passed before gains on the TOEIC can be detected (Agawa, Black, & Herriman, 2011). The online vocabulary program used in this study gave the students exposure to slightly less than 400 vocabulary words, but not all of the words may have been new to the students, and the actual amount of new vocabulary learned may have been, on average, considerably less. Had the study continued for another semester, it is possible that this form of vocabulary study would produce clearer advantages. A similar argument could be made for SSR and even explicit TOEIC practice and instruction, however. Only a longer-term study could let us know which practices are most effective in the long run.

Although a study which tracked student progress over a longer period of time would have likely clarified the effects of the different experimental approaches, it is noteworthy that despite the short duration of this study, students in all groups made considerable gains (100-150 points on average) on total TOEIC scores. Clearly these classes were helping the students improve within a 4-month time period, contrary to the somewhat disappointing results of earlier studies such as Robb and Ercanback (1999, see Section II-I).

In addition to the short time period, there are other weaknesses of this study that should be acknowledged. The differences between some of the groups on the pre-test mean scores prevent us from making strong conclusions about the results. Despite the utility of an ANCOVA to account for an initial inequality in mean scores, it is far more preferable to have a study in which all groups start at an equivalent level. Group 1 (+Voc) had the lowest gains in overall TOEIC scores and the reading portion, and nearly the lowest gains on the listening portion of the test, but this group also had the highest pre-test scores in all categories. It is possible that the content of the TOEIC courses was not ideally targeted for the higher level students, resulting in a less opportunity for progress. However, it should be acknowledged that Group 2 (+SSR, +Voc) had only slightly lower mean scores on the pre-tests, yet made the most progress over the course.

Another limitation of the study is that several groups in the study had different instructors, and though the instructors may have used identical materials and followed a similar format in conducting their classes, there are inevitably many subtle differences in teaching styles that may have an effect on student outcomes.

Despite these shortcomings of the study, the results of this study do show that time taken from explicit practice and instruction and given to SSR did not put the students at
any disadvantage. Groups 2 and 3 had roughly five fewer hours of explicit classroom instruction and practice than Groups 1 and 4, and yet they made equal or greater gains. Indeed, in the reading portion, both of the +SSR groups made superior gains on the reading portion of the test in comparison to the groups which had the full two hours of explicit study. This finding adds further support to findings in the general literature that SSR and extensive reading practices lead to equal, if not greater, gains than traditional teacher-centered instruction.

The results of the study can provide guidance for TOEIC instructors. Allowing time for SSR in class (or extensive reading as homework) can be an effective and enjoyable way for students to increase their overall reading speed. In the follow-up interviews, Professor L, who taught both of the groups using SSR in class, noted that the students seemed to look forward to the activity, and that a few students who tended to be inattentive during the explicit instruction and practice became far more active during the silent reading.

In regards to CALL vocabulary, a survey of the students conducted on the website found that most (82%) found this type of homework useful or very useful. As noted in Section II-3, the advantage of using online vocabulary programs is that they can be done as homework, freeing the instructor to utilize class time for other activities. There are many such programs available on the Internet. Many are free, though those which allow teachers to monitor student progress usually charge a fee.

Finally, it should be restated that both supplemental activities tested in this study, SSR and CALL vocabulary study, likely provide language learning benefits that extend beyond an increase on TOEIC scores. Explicit study of TOEIC may provide some benefit to language mastery beyond the test, but as most of the practices are so specific to test strategies and rather specific types of reading and listening texts, it is unlikely that general language skills will be markedly affected. As noted in the literature review, SSR and explicit vocabulary study extend well to overall language ability, both in the classroom and out. It is important for educators and students to keep in mind that the purpose of learning the global language of English extends far beyond any test score, and our classroom practices should reflect this as much as possible. Any activity which leads to gains on TOEIC that are equivalent to those received from explicit test practice, yet also develops skills beyond test taking, should receive serious consideration for adoption.

REFERENCES


proceedings/2009/E106.pdf


Pan, Y. C. (2010). Enhancing students’ communicative competency and test-taking skills
through TOEIC preparatory materials. TESOL Journal, 3, 81-91.


APPENDIX A
Student Survey

Name: _______ Student Number: ____________

1. 토익 공부를 1년간 (2011) 어떻게 해왔나요?

A. 토익 수업(예, 대구한의대 토익 수업, 온라인토익수업, 학원 등등..)을 구체적으로 쓰시오. ________________

B. 개인적으로 토익공부를 얼마나 했나요?
   ① 거의 한적이 없다
   ② 가끔씩 한다
   ③ 조금씩 정기적으로 한다.
   ④ 일정한 시간 매일 공부를 한다.

2. 당신은 정시 토익 응시를 한적이 있나요? (Yes No) 있다면 몇 번 응시를 했나요? ________

3. 여름 방학 동안, 당신은 영어 학습을 위해 무엇을 했나요? (학원수업, 개인공부, 토익공부 동)
당신은 여름방학 동안 주당 몇 시간 영어를 공부 했나요?

4. 토익 외에, 어떤 영어수업을 이번 학기에 들고 있나요? (Yes No)
   (영어로 진행 하는 수업도 쓰세요.)
   __________
   __________

5. 이번 학기 동안 학교 수업 외에 영어 수업 듣는 곳이 있나요? (Yes No)
   (학원, 개인공부 동)
   __________
   __________

6. 일반적으로 영어공부에 대한 본인의 생각은 어떻게시니까?
   ___ 영어공부를 좋아하지 않는다. 단지, 학점취득을 위해 공부한다.
   ___ 영어공부를 별로 좋아하지 않지만, 그냥 공부하는 편이다.
   ___ 영어를 좋아하지도 싫어하지도 않는다.
   ___ 영어공부를 조금 좋아한다.
APPENDIX B
Student Survey (English translation)

Name: ___________________ Student Number: __________________

1. What kind of TOEIC study have you done this year (2011)? (check all that apply)
   ___ TOEIC class (if the class is focused on only one aspect of TOEIC, such as listening only, please indicate this):
      For how long? ______________________
   ___ Private TOEIC study (if possible, write the name of your TOEIC books).
      For how long? ______________________

2. Have you taken an official TOEIC test before?
   If yes, how many times? ___

3. Over the summer break, what did you do for English study? (institute classes, private study, TOEIC study, etc.)
   About how many hours per week did you study English during the summer?

4. What other English classes are you taking this semester?
   ____________
   ____________
   ____________

5. What other English study are you doing these days (institutes, private study, etc.)
   ____________
   ____________
   ____________

6. What do you think about studying English?
   ___ I don’t really like studying English. I just want a good score on TOEIC to graduate or get a job.
   ___ I somewhat like studying English.
   ___ I enjoy studying English. I want to learn English well for reasons beyond just work and
study.

APPENDIX C
Reading Report

Name of book: __________________ Date: __________

How difficult was the book?

A. Very difficult. I need a dictionary to understand it.
B. A little difficult. Sometimes I can’t understand, but sometimes I am OK.
C. OK. Usually I can understand it.
D. Easy. I understood almost everything with no problem.

How enjoyable was the book?

A. Very enjoyable.
B. Somewhat enjoyable.
C. Just OK.
D. A little boring
E. Very boring

Response to the book: Write 6-8 sentences about the book (in Korean)
(brief synopsis, what information was most interesting, general opinion of the book)

Applicable Levels: Low intermediate-advanced
Key Words: vocabulary learning, CALL, silent sustained reading, TOEIC preparation

Scott W. Miles
Department of English TESOL
Daegu Haany University
9-314, Yugok-dong
Gyoungsan, North Gyoungsan Province, 712-715,

Email: scott@dhu.ac.kr
Received in June, 2012
Reviewed in July, 2012
Revised version received in August, 2012