

Enhancing Reading Comprehension and Summarization Abilities of EFL Learners Through Online Summarization Practice

Chia-Hui Chiu¹

Abstract

This study investigated the effects of online summarization practice on the reading comprehension and summarization abilities of EFL learners. Thirty-five EFL learners read and summarized expository passages using an online tutoring program called WriteToLearn. Upon online submission of their summaries, the learners received instant automated feedback and had multiple opportunities to revise summaries. Another group of 34 learners serving as a comparison group read the same passages and did comprehension exercises which required them to provide answers to comprehension questions in writing. Both groups took a reading comprehension test and a summarization test once before and once after the four-week training. The results showed that while the two groups exhibited significant improvement in micro-level and macro-level reading comprehension and summarization, the online summarization group performed significantly better than the comprehension exercise group in macro-level reading comprehension and summarization after training. The findings suggest that online summarization practice is superior to the traditional procedure of answering comprehension questions for helping learners enhance reading comprehension and summarization abilities.

Key words: Reading comprehension, Summarization, Online tutoring program

1. Introduction

Summarization involves identifying main ideas, deleting unimportant information, and reconstructing ideas from original texts concisely. It is seen as a useful metacognitive strategy that can monitor and facilitate comprehension and ultimately contributes to better-integrated memory representation of the knowledge base (Kintsch, 1988, 1998). Since the 1990s, L1 educational research has gathered ample evidence that summarization is an effective strategy for deeply understanding reading materials (Bean & Steenwyk, 1984; Britt & Sommer, 2004; Jitendra, Cole, Hoppes, & Wilson, 1998; Jitendra, Hoppes, & Xin, 2000; Malone & Mastropieri, 1992; Radmacher & Latosi-Sawin, 1995; Rinehart, Stahl, & Erikson, 1986; Rogevich & Perin, 2008; Trabasso & Bouchard, 2002; Westby, Culatta, Lawrence, & Hall-Kenyon, 2010). Although the task of summarizing in L2 is more taxing for L2

¹ Tunghai University, Taiwan, Email: chiahuichiu@thu.edu.tw

learners (Keck, 2006; Kim, 2001), L2 research has also found substantial effects of summarization on the reading comprehension of EFL learners (Baleghizadeh & Babapur, 2011; Bensoussan & Kreindler, 1990; Cordero-Ponce, 2000; Oded & Walters, 2001; Shokrpour, Sadeghi, & Seddigh, 2013).

Recently, the development of computer and educational technologies has strongly and positively influenced language teaching and learning. Research findings have typically shown that learners can achieve significant learning gains through computer-assisted language learning (CALL), and generally view CALL positively (Fotos & Browne, 2004). Given the facilitating effects of summarization and the great promises of CALL, this study examines the extent to which summarization practice provided by an online tutoring program called WriteToLearn benefits EFL learners.

2. Related Literature

2.1 The Construction-Integration Model

The Construction-Integration model, proposed by van Dijk and Kintsch (1983) and later extended by Kintsch (1988, 1998), has been most influential in reading research. According to this model, reading comprehension involves an initial phase of construction followed by an integration process. During the construction process, the reader constructs a network of propositions formed via the analysis of the linguistic input and/or generated from the prior background knowledge of readers. The process is conducted cyclically. During each cycle, phrases or sentences are processed and a small number of propositions are formed and retained in working memory to be processed with the material from the next cycle. During this phase, the text representation is most likely to be carelessly constructed and largely based on the overlap among propositions. Text representation may be incoherent because it may include unimportant propositions unrelated to the theme of the discourse context. During the next phase, integration is performed to exclude the unwanted propositions from the text representation based on the constraints imposed by the text and the prior knowledge of the reader. During this process, each proposition gains a renewed level of activation that may differ from its initial activation. Irrelevant propositions formed in the construction phase are deactivated, whereas propositions that receive higher activation values are retained. A coherent text representation is then formed following the integration process.

The text processing is accompanied by the establishment of two types of propositions, micro and macro. Micropropositions represent meaning at the sentence level of a text and create a local understanding of the text at the microstructure level. Macropropositions organize the micropropositions according to their relative relevance to the text theme and contain summary propositions at the macrostructure level, and in so doing denote the global structure of the text. Although both types of propositions are important, the macropropositions contribute more than the micropropositions to overall text comprehension and long-term knowledge accumulation (Kintsch, 1988).

2.2 Effects of Summarization

Reading for academic purposes is commonly combined with writing activities such as note-taking, answering post-reading questions, summarizing and writing response papers. Among these types of writing, summarization has attracted the most research attention because it is believed to be the kind of strategy instruction that helps learners comprise macropropositions of a text and hence better understand the text. When summarizing a text, learners engage in a process that involve identifying the main ideas in a passage, distinguishing major and minor details, and restating the main ideas with the goal of expressing the gist of the passage.

Since the 1990s, L1 researchers have conducted studies in experimental settings to investigate the extent to which summarization helps learners improve reading comprehension (e.g., Britt & Sommer, 2004; Radmacher & Latosi-Sawin, 1995; Rinehart, Stahl, & Erikson, 1986; Westby, Culatta, Lawrence, & Hall-Kenyon, 2010). Such studies tend to focus on summarizing expository texts because they present more conceptual difficulties for classroom learners than narrative texts. For instance, in a study conducted by Radmacher and Latosi-Sawin (1995), 16 American university students read different summaries of a passage, discussed their quality and then summarized one section selected from the psychology textbook weekly. After four weeks of training, their scores in the text comprehension exam were compared with the scores of 17 students who simply listened to lectures on the same sections of the textbook. The results showed that while the first exam mean score of the summarization group was 7% lower than that of the control group, their final exam mean score was 8% higher, and thus statistically differed from that of the control group. Experimental studies carried out at the elementary school level also reported direct and indirect effects of the summarization program on reading and studying skills for fourth- and fifth-grade students (Westby, Culatta, Lawrence, & Hall-Kenyon, 2010) as well as sixth-grade students (Rinehart, Stahl, & Erikson, 1986).

The same facilitative effects of summarization practice have also been reported in studies involving L1 learners with learning difficulties associated with behavior or attention disorders (Jitendra, Cole, Hoppes, & Wilson, 1998; Jitendra, Hoppes, & Xin, 2000; Malone & Mastropieri, 1992; Rogevich & Perin, 2008). Rogevich and Perin (2008) studied 63 adolescent boys with behavioral disorders who read science texts and investigated whether these learners could benefit from an intervention called Think before reading, While reading, and After reading with Written Summarization (TWA-WS). This intervention stressed self-monitoring and comprised several steps in which learners identified the purpose of the author, set a reading goal, linked their background knowledge to new information in the text, identified the main ideas, summarized the important information from the text and reflected on what they had learned. The results demonstrated that after completing five training sessions, TWA-WS participants exhibited significantly greater gains than a matched comparison group that received traditional literacy practice on reading comprehension as measured using written summarization.

Summarization has attracted considerable attention from L2/EFL researchers as well and they have reported positive effects of summarization (Baleghizadeh & Babapur, 2011; Bensoussan & Kreindler, 1990; Cordero-Ponce, 2000; Oded & Walters, 2001; Shokrpour, Sadeghi, & Seddigh, 2013). Employing an experimental-control-group design, Cordero-Ponce (2000) studied the effects of explicit summarization instruction in a French course that involved breaking down complex skills, modeling and guided and independent practice. The results indicated that the experimental group significantly outperformed the control group in reading comprehension and summarization after completing the training. Baleghizadeh and Babapur (2011) and Shokrpour, Sadeghi and Seddigh (2013) conducted their studies with university EFL students in the Middle East and obtained the same findings. However, neither study detailed how summarization was taught, nor were the differences between the treatments for the experimental and control groups made clear. Lack of such information makes the interpretation of the results less clear.

A few earlier studies compared the effects of summarization on reading comprehension with other tasks that require different degrees of reading-related processing. For instance, Bensoussan and Kreindler (1990) completed a one-semester reading course at Haifa University that examined the effects of summarization and answering short-answer questions. A group of 92 learners summarized academic texts of general interest and discussed summarization strategies. The other group of 87 learners completed short-answer questions on the same texts and discussed strategies for correctly answering the questions. The results showed that after training both groups improved significantly on the reading test, and the two groups appeared not to differ significantly. The researchers claimed that the parallel improvement displayed by the two groups could be attributed to classroom discussion, in

which students belonging to both groups touched on both macro-level and micro-level propositions. Therefore the overlapping components of the two treatments probably made it difficult to differentiate the effects of summarizing and answering short-answer questions on reading comprehension. Another possible explanation for the result was that the reading posttest mostly tested micro-level rather than macro-level ideas, and hence was biased toward the question answering group. In a well-controlled study, Oded and Walters (2001) compared the effects of summarization and details listing. Sixty-five EFL learners at an Israeli university participated in two sessions of a reading activity. In the first session, half were asked to list all the examples appearing in a text, while the other half were asked to summarize the same text in writing. In the second session, those who listed examples in the first session summarized another text. Those who summarized in the first session listed examples. Multiple-choice comprehension questions were prepared for the two texts used in the two sessions. The questions focused on main ideas, the purpose of the writing, the organization of the text and key supporting evidence. The results confirmed the researchers' hypothesis that the depth of processing required in selecting the main ideas and organizing them in the summarization task increased comprehension; the listing of details, being irrelevant to overall comprehension, resulted in poorer text comprehension.

One of the most important aspects of summarization is the ability to paraphrase. Good paraphrasing ability shows that the learners have understood the articles and can use the authors' main ideas in their own way. Unlike the L1 context, in which copying from the source texts during summarization does not appear to be a concern, L2 researchers have explicitly noted the need to teach L2 learners how to paraphrase and restate the main ideas of the source text in their own words (Keck, 2006; Kim, 2001). In a study comparing the use of paraphrasing in summarization among L1 and L2 college students, Keck (2006) classified the paraphrases of participants into four types: near copy, minimal revision, moderate revision, and substantial revision. She found that L2 learners used significantly more near copies than L1 learners, while L1 learners displayed significantly more moderate and substantial revisions than L2 learners in their summarizations. Accordingly, a few L2 researchers placed additional emphasis on paraphrasing strategies when teaching summarization. For example, Choy and Lee (2012) taught summarization skills to 22 Malaysian EFL learners enrolled in a two-year diploma program for 10 weeks. Throughout the summarization process, the learners were told not to copy the original sentences from the source text and instructed on how to perform word and phrase substitution. The results demonstrated that strategy instruction on paraphrasing did not help all learners equally. Only 36% of learners, including those who perceived themselves as having benefited from the instruction, exhibited improved performance on the summarization posttest. In the study conducted by McDonough, Crawford and De Vleeschauwer (2014), 46 Thai EFL non-English major university students underwent training on how to write paragraphs and use paraphrasing strategies for summarization during a 17-week EFL writing class. The results demonstrated significant decreases in the occurrence of copied word strings and increases in the occurrence of modified word strings among the learners, suggesting the paraphrasing instruction was effective. One possible explanation for the differences in findings between McDonough et al. (2014) and Choy and Lee (2012) may relate to the English proficiency of the learners. The university students in the McDonough et al. study (2014) most likely had better English ability than those in the Choy and Lee study (2012) who were only enrolled in a two-year diploma program. Having a better command of English, university students may have benefited more from the paraphrasing strategy instruction and thus showed significant improvement in summarization.

In sum, summarization of texts can improve reading comprehension and has significantly better effects than other reading-related learning activities. Summarization, however, is not easy for L2 learners. In addition to identifying the main ideas, deleting unimportant details, and generalizing during reading, many L2 learners need adequate guidance to restate the points of the author in their

own words. Any L2 summarization instruction should take all these subskill limitations into consideration.

2.3 Latent Semantic Analysis: Basis for Online Essay Assessment

Despite the extensive evidence of the positive effects of summarization, summarization does not appear to be commonly taught in regular classrooms and practiced outside of classrooms (Landauer, Lochbaum, & Dooley, 2009; Ono, 2011). This is so because assessing summarization and providing feedback may create a heavy workload for teachers, particularly when numbers of students and/or size of summaries are large, which causes delays in providing feedback (Flotz, Gilliam, & Kendall, 2000; Landauer, Lochbaum, & Dooley, 2009). Recent developments in CALL tutors can potentially be of help in this situation because they are designed to provide learners with opportunities for extensive and repeated practice in specific language skills and, most importantly, assess learner performance online and offer instant automated feedback (Levy & Stockwell, 2006: 22).

An important basis for such online assessment is Latent Semantic Analysis (LSA), a mathematical and statistical technique used to construct knowledge representation in the form of a high-dimensional semantic space (Kintsch, Ericsson, & Patel, 1999). Its basic assumption is that every document has an underlying semantic structure that can be quantified in a matrix. In the operation of LSA, the computer reads as input a large amount of text comprising thousands of documents and tens of thousands of words. Based on this input, LSA can construct a large word-by-context matrix, in which each row represents a unique word and each column represents a passage, paragraph, or sentence of text. The entries to each cell are the frequency counts for the number of times each word appears in each document. The input is then, as described by Kintsch et al. (1999) and Landauer, Foltz and Laham (1998), processed by weighting each cell frequency via a function that expresses both the importance of the word in that particular document and the degree to which the word carries information related to that specific domain of discourse. However, this original matrix is not desirable for knowledge representation because it contains too much information. The use of specific words in particular documents is unimportant. Rather what matters is the kinds of words that could be used in that particular document. That is, LSA concerns the meaning relationship between words and context, not word choices. Thus, in the final step, termed dimension reduction, singular value decomposition is applied to the matrix, which allows it to be expressed as the product of three matrices, one of which is the singular values matrix. In this step, LSA retains the information associated with the roughly 400 largest singular values of the matrix and discards the incidental and remote details. Wade-Stein and Kintsch (2004) asserted that when a high-dimensional semantic space of 300-400 dimensions is constructed, LSA can express words, sentences, paragraphs and whole texts as vectors, and can readily compute the semantic relatedness of vectors in terms of the cosine, which can be interpreted as a correlation. A large cosine means that the texts contain similar meanings and are located in close proximity to one another within the space, or that the words occur in similar contexts and are grouped together as semantically related in the space.

2.4 Application of Latent Semantic Analysis

Although LSA was not developed as an educational tool, its ability to judge semantic relatedness is readily applicable to essay scoring. Essays require learners to apply their knowledge to write about a field so in this sense, essays also assess learner knowledge of a specific area, where this learner knowledge primarily derives from reading texts. Thus, the degree of similarity in meaning between the essay and text should be a good indicator of essay quality (Miller, 2003). Empirical studies such as Landauer, Laham, Rehder and Schreiner (1997) and Flotz, Gilliam and Kendall (2000) reported correlations between LSA measure and human scores of up to 0.77, similar to the intercorrelations of

scores provided by professional graders. Flotz et al. (2000) found that students enrolled in an undergraduate course achieved significant progress in writing on a topic through revising their essays in response to LSA-based system-generated feedback.

LSA has been applied to several commercial online tutoring programs, such as WriteToLearn, in which students can enter their essays or article summaries and a few seconds following the online submission, receive an estimated grade together with suggestions for revisions. When LSA is applied to summarization assessment, a cosine can be calculated between the summary written by a student and the source text. If the cosine is below a certain empirically determined threshold value for content coverage, WriteToLearn informs the learners that their summary does not yet cover the content sufficiently. The redundancy and relevance checks inform the learners that some portions of their summaries may be redundant and/or unimportant to the main ideas of the source text. The copy check informs the learners that some part of their summaries may be directly copied from the source text. The feedback on copying is deemed to be helpful scaffolding for weak learners who have problems improving their paraphrasing skills. Finally, the spelling check informs the learners regarding misspelled words. In short, the types of automated feedback provided by WriteToLearn mostly correspond to the key components of a good summary, and hence have the potential to facilitate the development of important subskills of summarization. Overall, the immediacy of the feedback holds great promise to motivate learners to revise summaries while they still remember the source texts.

Several researchers have carried out classroom trials to study the effects of the online summarization practice provided by the precursor of WriteToLearn, called Summary Street (e.g., Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005; Wade-Stein & Kintsch, 2004). Wade-Stein and Kintsch (2004) had 47 sixth-grade students summarize articles using Summary Street and receive either content-based or other forms of feedback. Human raters blindly scored the student summaries. The researchers found that learners who received content-based feedback performed significantly better and spent more than twice as long on the summarization task compared to those who used a visually similar interface that provided only information regarding spelling errors and vertical length indicators but no content feedback. Their results suggested that online summarization practice with instant automated content-based feedback effectively directs students to pay attention to relevant content. Similarly, Franzke, Kintsch, Caccamise, Johnson and Dooley (2005) compared the learning gains of 111 eighth-grade students, who practiced summarizing over a 4-week period, either with or without the guidance of Summary Street. Blind scoring indicated that the student summaries created using Summary Street were significantly superior on several measures of writing quality, including overall quality, more complete content coverage, better organization, lack of unnecessary detail and good stylistic quality. The researchers also found that the experimental group significantly outscored the control group on test items that measured reading comprehension at the macro level. These studies have shown that Summary Street can get learners to closely engage in online summarization through instant, interactive and content-based iterations of the online program and effectively improve their comprehension and summarization abilities.

3. The Present Study

The practical values of online summarization practice, however, are not clear unless comparisons are made to other commonplace classroom activities or traditional homework assignments such as answering short-answer comprehension questions. An investigation of this issue is important before one can claim that summarization practice using online technology is a better choice than traditional methods for improving reading comprehension and summarization abilities of students in a school setting.

The present study intends to contribute to this body of research by comparing the effects of online summarization and answering comprehension questions. The activities of answering comprehension questions and summarizing are closely related because both require learners to utilize the same strategies such as reading selectively and carefully in key places, rereading as appropriate, identifying key information, and integrating ideas from different parts of the text. However, summarizing differs from answering questions in degree of processing complexity. When summarizing, learners are likely to consciously monitor their reading and manipulate their linguistic knowledge in a productive way, through which a sophisticated process of meaning construction in writing is carried out (Bensoussan & Kreindler, 1990). Further, when summarization is conducted in the online WriteToLearn program, the instant automated feedback can provide additional help for learners to refine the meaning construction involved in summarization. In contrast, answering comprehension questions are prepared to help learners to relate text information to the types of questions posed. Learners are invited by the questions to pay attention to and search for directly available information in the text or make inferences based on information supplied or draw upon their background knowledge (Grabe, 2009: 232). When answering comprehension questions, learners do not need to engage themselves in the complicated process of meaning construction. It is hence hypothesized in the present study that online summarization practice leads to larger learning gains relative to the traditional answering of comprehension questions.

A quasi-experimental design was employed in this study. One group of learners performed online summarization practice using WriteToLearn and thus was labeled the online summarization group. Meanwhile, another group conducted comprehension exercises and thus was labeled the comprehension exercise group. The reading comprehension and summarization abilities of the two groups were examined and compared after the experiment. This study sought to answer the following research questions:

1. Does any significant difference exist between the online summarization group and the comprehension exercise group in micro-level reading, macro-level reading and summarization upon completion of the four-week training?
2. Do the two groups exhibit significant improvements in micro-level reading, macro-level reading and summarization upon completion of the four-week training?

4. Method

This section details the participants, overview of the study design, instructional treatments, and instruments used in this study. The procedure for data collection and data analysis are also described.

4.1 Participants

Sixty-nine EFL learners recruited from two Freshman English classes at a private university in central Taiwan participated in this study. Their English proficiency, as indicated by performances on the English placement examination developed by the university, was ranked among the top 10% of all freshmen enrolled the same year in the university. One class comprised 35 students majoring in management, including 19 females and 16 males, while the other comprised 34 students majoring in social science, including 20 females and 14 males. The management class was randomly assigned to the online summarization group and the social science class the comprehension exercise group.

4.2 Overview of the Study Design

Freshman English at the research site was a year-long course. Each class met three times a week for five 50-minute class periods. In the fall semester, the researcher/instructor taught learners strategies essential for successful reading comprehension, including predicting the content of the article through reading titles, headings, and pictures; using context clues to guess the meanings of unknown words; and identifying the key point of an article by locating and distinguishing major and minor details. The study was conducted in the spring semester of year 2010. During the first three weeks, the learners gave oral presentations sharing their recreational activities during the winter break, took the reading and summarization pretests and attended a pre-training workshop that dealt specifically with the type of homework the learners would engage in. From Week 4 through Week 7, the online summarization group read passages selected from the online WriteToLearn program and performed summarization exercises. The comprehension exercise group read the same passages but conducted comprehension exercises. The researcher/instructor spent the same amount of time providing both groups instruction regarding the different learning tasks they performed. In Week 8, both groups took the reading and summarization posttests.

4.3 Instructional Treatments

Pre-training workshops. Pre-training workshops with different topics were held for the two groups in Week 3. During the 50-minute workshop, the WriteToLearn program was introduced to the online summarization group in a computer lab, where the learners familiarized themselves with the interface of the online program and learned how to use the editing tools for summarizing. Specifically, the online summarization group was taught that a good summarizer should write a clear thesis statement at the start to provide a brief overview of the substance of a passage, and then include important main points of the passage. It was emphasized that a good summarizer generally restates the main ideas using their own words. The learners then learned to use the function keys to submit summaries to the online system to receive automated feedback. Figure 1 shows a sample feedback window that was demonstrated to the learners. The researcher/instructor explained that the horizontal bars on the upper left corner of the window give an approximation of how well the summary covers the content of each section of the passage. Similarly, the vertical bar on the upper right corner gives an approximation of the length of the summary. The bars move on a theoretical continuum from the red zone (indicating poor performance), through the yellow zone (indicating fair performance), and finally toward the green zone (indicating excellent performance). The percentages of the content of the summary that contains unwanted behaviors of summarization are shown in the lower part of the feedback window. The unwanted behaviors refer to copying directly from the original source text, making spelling mistakes, and including repetitive, irrelevant and unimportant information in the summary. The percentages are highlighted using a color coding system, with red indicating that 20% or more of the summary contains unwanted behaviors, yellow indicating 5% to 20%, and green indicating less than 5%. The workshop concluded by having the online summarization group work together to summarize a short passage and submit the summary to the online system. To show the learners how to revise the summary according to instant automated feedback, the researcher/instructor clicked on the sentences highlighted in red or yellow and made efforts to make additions or deletions, or rephrase the entire sentence.

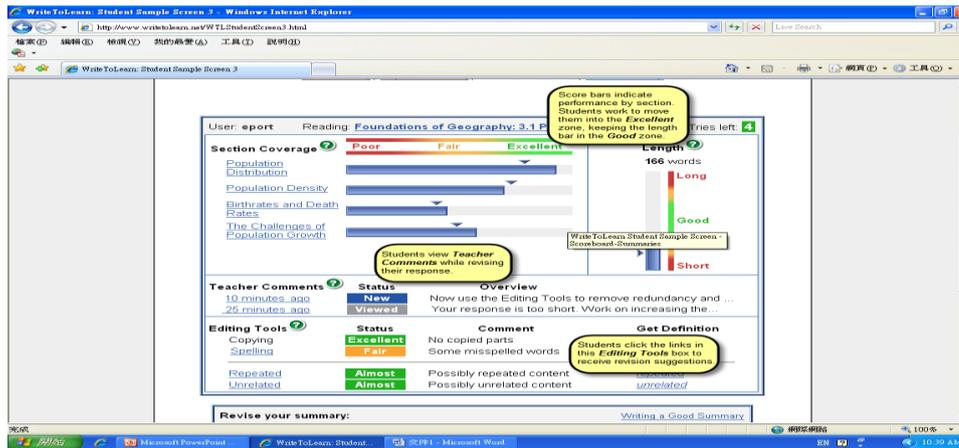


Figure 1. Sample Feedback Window of WriteToLearn

For the comprehension exercise group, the researcher/instructor gave an oral presentation providing information on different language tests (GEPT, TOEFL, IELTS and TOEIC) that Taiwanese EFL learners may be interested in taking or required to take before graduating from college. The purposes and different sections of the tests were briefly reviewed with special attention paid to the reading section of the TOEFL, and various types of multiple-choice questions (e.g. best title, main ideas, facts, vocabulary, and references) were analyzed. The workshop concluded by having the comprehension exercise group read sample TOEFL articles and complete reading comprehension questions that required the EFL learners to write short answers in response.

Instructional passages. From Week 4 through Week 7, the learners in both groups received a hard copy of an expository passage each Monday. Passage selection was crucial in this study because WriteToLearn was originally designed for use in the L1 setting and the passages labeled suitable for American students at each grade level might not be suitable for the EFL learners participating in this study. Thus, during the piloting stage, eight passages considered appropriate for the EFL learners were pre-selected. The interests and English proficiency level of the learners were taken into account during the selection process. The eight passages were then piloted with a group of 20 EFL learners from another high-level class not participating in the study. The 20 learners were asked to rate the difficulty of the eight passages using a 5-point scale (1 = very easy to understand, 2 = a bit easy to understand, 3 = average, 4 = a bit difficult to understand, and 5 = very difficult to understand). Of the eight pre-selected passages, six received average ratings ranging between 3.85 and 4.10, and the last two received average ratings of 4.35 and 4.50, respectively. An informal group interview revealed that although the topics of the eight passages did not require specific background knowledge, the two passages with the highest average ratings were considered too difficult owing to a high percentage of unknown or unfamiliar vocabulary items. To keep the learners motivated to read and not frustrated by the large number of unknown or unfamiliar English words, these two passages were excluded. Four of the six remaining passages were randomly selected as instructional texts during training. The titles of these four passages were "Globalization," "Advances in Science and Technology," "City Life" and "The Art of Making Mummies." The other two passages served as articles to be summarized during the summarization pretest and posttest. The titles of the two passages were "Nation on the Nile" and "China: Transforming Itself." The passages were organized via a mixture of time-order and description, comprised four to seven sections and were approximately 1,100 words long. These passages were marked by the program designer as suitable for 6th - 8th grade American students.

Homework assignments. The EFL learners received copies of the passages on Monday to complete homework assignments due on the Friday of the same week. The online summarization group was required to summarize the passages using the WriteToLearn program, encouraged to

revise their summaries upon receiving automated feedback, and permitted to revise their summaries up to five times. The researcher/instructor discussed the automated feedback with the EFL learners the following week. Meanwhile, the comprehension exercise group was asked to answer reading comprehension questions related to the passages they read. The question types resembled those in the TOEFL test. However, unlike the TOEFL test, which uses a multiple-choice format, the assignments required learners to answer questions by writing complete sentences. The researcher/instructor graded their assignments and reviewed the correct answers with the learners the following week. The EFL learners were asked to resubmit their corrections for checking.

During the four-week training, except for the different assignments administered to the online summarization and the comprehension exercise groups, the classroom instruction and activities for the two groups were identical and were designed to offer learners opportunities to practice speaking and communicative skills. That is, other in-class activities were controlled to ensure that after the training, any significant differences between the two groups in terms of their reading comprehension and summarization performances would be attributed to their different training.

4.4 Instruments

Reading tests and summarization tests were designed to measure the reading comprehension and summarization abilities of the EFL learners before and after the training.

Reading test. Two different but equivalent forms of the reading test (Forms A and B) were created. The passages and multiple-choice questions were adapted from the reading section of retired TOEFL tests. Each form contained three reading passages, which were organized according to a mixture of time order and description and required no specialist knowledge to understand. Each passage comprised approximately 450-500 words and contained ten questions. Five were macro-level questions testing comprehension of the main idea of the paragraph(s), or the entire passage and inferences that could be made based on the passage. The other five were micro-level questions testing the meanings of words and the stated facts of the passage. Each question was worth one point. A maximum of 30 points could be earned on the reading test, with 15 from micro-level questions and a further 15 from macro-level questions. To ensure the two forms were of comparable difficulty, the 20 EFL learners who helped with the selection of instruction texts also volunteered to answer the questions on the two forms during the pilot-testing session. To counterbalance the sequence effect, 10 learners were randomly selected to complete Form A followed by Form B, while the other 10 learners completed the two forms in the reverse order. The mean scores of these volunteers for the micro-level questions were 5.60 on Form A and 5.55 on Form B, while for the macro-level questions the mean scores were 5.75 on Form A and 5.60 on Form B. The results of paired t-tests indicated no significant differences in their performances on the micro-level questions of the two forms ($t = 0.108$, $p = .915$) and on the macro-level questions of the two forms ($t = 0.513$, $p = .614$). The two forms thus could be considered comparable in terms of both the micro- and macro-level questions.

Summarization test. The two passages, titled "Nation on the Nile" and "China: Transforming Itself", with neither having appeared in instructional texts, were given to the learners to summarize as the summarization pretest and posttest, respectively. It was stated in the test instructions that a summary should start with a thesis statement to indicate the central theme of the passage, and include only the authors' main ideas. The learners were also required to restate the main ideas in their own words. In addition, it was recommended that the summary be one fifth to one fourth as long as the original source text. The researcher/instructor and a native English speaker teaching English in a different college assessed the summaries using a scoring rubric modified from Friend (2001). Each summary was scored for four components of summarization: (a) thesis statement, (b) content inclusion and exclusion, (c) sentence transformation, and (d) grammar and mechanics (see Appendix 1). The second and third components are key aspects of the summarization in WriteToLearn.

Although the first and fourth components are not assessed in the online program, the two components are considered essential to summarization (Friend, 2001). The two raters independently assigned scores for each summary. The maximum score for each aspect was 4 while the minimum was 1. Thus, the maximum total score for the summarization test was 16 and the minimum was 4. When the difference between the total scores given by the two raters exceeded 2 points, the discrepancy was resolved through discussion. The final score for each summary was the average of the two scores given by the two raters after such discussion.

4.5 Data Collection and Analysis

The reading and summarization pretests were administered to both groups in class during Week 2. The reading and summarization posttests were administered during Week 8. The two research questions concern the differences between the two treatment groups and the improvement achieved by each group in micro-level reading, macro-level reading and summarization upon completion of the training. A series of hypothesis testing were performed to compare the mean scores with a significance level of $\alpha = .05$. MANOVA, which is considered appropriate for comparing multiple mean scores, was employed.

5. Results

Table 1 lists the means and standard deviations of the test scores of the EFL learners. As the table shows, the online summarization group displayed improvement in the micro-level (pretest score mean = 5.83 vs. posttest score mean = 7.17) and macro-level reading questions (pretest score mean = 6.26 vs. posttest score mean = 7.97). Similarly, the comprehension exercise group also exhibited improvement in the micro-level (pretest score mean = 5.91 vs. posttest score mean = 7.00) and macro-level reading questions (pretest score mean = 6.29 vs. posttest score mean = 6.97). Meanwhile, both groups showed improvement for summarization. The mean scores of the online summarization group were 9.23 in the pretest and 10.71 in the posttest. The mean scores of the comprehension exercise group were 9.16 on the pretest and 9.93 on the posttest. The performances of the two treatment groups were similar in the pretests but differed somewhat in the posttests.

Table 1. Means and Standard Deviations of Test Scores of EFL Learners

Group	Pretest			Posttest		
	Micro-level Reading ^a	Macro-level Reading ^b	Summarization ^c	Micro-level Reading ^a	Macro-level Reading ^b	Summarization ^c
Online Summarization Group (n=35)	5.83 (2.15)	6.26 (2.72)	9.23 (1.57)	7.17 (1.79)	7.97 (1.71)	10.71 (1.21)
Comprehension Exercise Group (n=34)	5.91 (2.15)	6.29 (2.53)	9.16 (1.49)	7.00 (1.86)	6.97 (2.28)	9.93 (1.62)

Note. ^aThe maximum score is 15. ^bThe maximum score is 15. ^cThe maximum score is 16.

Reliability indices were computed for the tests. Satisfactory reliabilities were achieved in the two reading tests, with Cronbach's $\alpha = .82$ for the pretest and $.84$ for the posttest. Inter-rater reliability indices were computed for the summarization tests. The inter-rater reliabilities were good for both the pretest ($r = .90$ before discussion and $r = .95$ after discussion) and posttest ($r = .91$ before discussion and

$r = .97$ after discussion). The following paragraphs present the results of hypothesis testing in the order of the two research questions.

RQ1: Between-Group Comparisons

Before examining whether significant differences in test scores existed between the two groups upon completion of the training, a MANOVA was performed to ensure the two groups were at a similar baseline before training. The learner's group was the independent variable, and its micro-level reading, macro-level reading and summarization pretest scores were the dependent variables. The multivariate test results demonstrate no statistically significant difference between the two groups in their pretest mean scores (Wilks' $\Lambda = 0.996$, $F(3, 65) = 0.097$, $p = .961$), implying the two groups could be considered equivalent at the outset of the training. Another MANOVA was then performed to examine whether any significant difference existed between the two groups in posttest mean scores. The multivariate test results demonstrate at least one significant difference between the two groups in the three posttest mean scores (Wilks' $\Lambda = 0.876$, $F(3, 65) = 2.279$, $p = .047$). Univariate tests were then performed to further identify the differences. The results show that the mean scores of the two groups differed significantly on the macro-level reading questions ($F(1, 67) = 4.288$, $p = .042$) and the summarization test ($F(1, 67) = 6.335$, $p = .014$) but not the micro-level reading questions ($F(1, 67) = 0.152$, $p = .698$), implying that upon completion of the training, the online summarization group significantly outperformed the comprehension exercise group in macro-level reading and summarization but not in micro-level reading.

RQ2: Within-Group Comparisons

To examine whether the two groups displayed significant improvement in micro-level and macro-level reading and summarization upon completion of the training, a repeated-measures MANOVA was performed for each group with time as the within-subjects variable and the micro-level reading, macro-level reading and summarization test scores as the dependent variables. The multivariate test results demonstrate that for the online summarization group, at least one significant difference existed between their pretest and posttest mean scores (Wilks' $\Lambda = 0.446$, $F(3, 32) = 13.256$, $p < .001$). Univariate tests were then performed to further identify the differences. The results show that for the online summarization group, the pretest and posttest mean scores differed significantly in the micro-level reading questions ($F(1, 34) = 6.545$, $p = .015$), macro-level reading questions ($F(1, 34) = 13.290$, $p = .001$) and summarization test ($F(1, 34) = 40.888$, $p < .001$). The same procedure was followed for the comprehension exercise group and yielded similar findings. The multivariate test results also indicate at least one significant difference between the pretest and posttest mean scores (Wilks' $\Lambda = 0.253$, $F(3, 31) = 30.512$, $p < .001$). The univariate test results show that the pretest and posttest scores differed significantly for the micro-level reading questions ($F(1, 33) = 38.253$, $p < .001$), macro-level reading questions ($F(1, 33) = 23.947$, $p < .001$) and the summarization test ($F(1, 33) = 28.382$, $p < .001$). In sum, both the online summarization and comprehension exercise groups displayed significant improvement in micro-level reading, macro-level reading, and summarization after four weeks of training.

6. Discussion and Conclusion

This study found that both the online summarization and comprehension exercise groups made significant improvements in reading comprehension and summarization after training. Although online summarization and answering comprehension questions are different activities, the similarity of learning gains for both groups suggests that both tasks had similar facilitating effects. Moreover, the findings of this study regarding the significantly better performance of the online summarization group compared to the comprehension exercise group are in line with those of previous research that

reported significant efficacy of summarization practice on improving learners' reading comprehension and summarization abilities, compared to other types of reading-related learning tasks (L1 studies: Radmacher & Latosi-Sawin, 1995; Rinehart et al., 1986; Rogevich & Perin, 2008; Westby et al., 2010; and L2 studies: Baleghizadeh & Babapur, 2011; Bensoussan & Kreindler, 1990; Cordero-Ponce, 2000; Oded & Walters, 2001; Shokrpour et al., 2013). Different from Bensoussan and Kreindler (1990), the present study further found that online summarization group performed significantly better than the comprehension exercise group on the macro-level reading questions. These findings confirm the researcher's hypothesis that the online summarization practice would lead to greater improvement, relative to the traditional answering of comprehension questions.

Similar to the findings reported by Wade-Stein and Kintsch (2004) and Franzke et al. (2005), the significantly better performance of the online summarization group may be due to the benefits gained from the instant automated feedback generated by the online tutoring program. Although learners in this group were not required to revise their summaries, a calculation of the numbers of learners who submitted revised drafts shows that many learners did revise their summaries at least once. Table 2 lists the numbers. It was highly likely that the online automated feedback effectively helped learners pinpoint problematic phrases or sentences, motivated them to re-read the main points of the passages, and further pushed them to think deeply regarding ways to improve their summaries. The engagement during the subsequent revision process in summarization consequently led to learner improvement in reading comprehension and summarization abilities.

Table 2. Numbers of Learners in Online Summarization Group Who Submitted First, Second, Third, and Fourth Drafts for Each Summary

	Summary 1	Summary 2	Summary 3	Summary 4
First draft	35	35	34	33
Second draft	30	31	24	25
Third draft	16	14	9	12
Fourth draft	3	1	0	0

To further explore which aspect of summarization the online summarization technology was most useful in capturing, the researcher examined and compared the changes in scores of the two groups of EFL learners from the summarization pretest to the posttest. Table 3 shows the score increases according to the four components of the summarization scoring rubric.

Table 3. Scores Increases in Different Aspects of Summarization

Group	Thesis statement	Content inclusion and exclusion	Sentence transformation	Grammar & mechanics
Online Summarization Group (n = 35)	0.30	0.50	0.46	0.23
Comprehension Exercise Group (n = 34)	0.04	0.31	0.21	0.21

To determine whether the learning gains of the online summarization group in the four components of summarization significantly exceeded those of the comprehension exercise group, independent t-tests were performed. The results demonstrate that no significant differences were observed in content inclusion and exclusion ($t = 1.340, p = .185$) or grammar and mechanics ($t = 0.186, p = .853$). The score increases of the online summarization group significantly exceeded those of the comprehension exercise group in thesis statement ($t = 2.007, p = .048$) and sentence transformation ($t =$

2.392, $p = .020$). The non-significant difference in content inclusion and exclusion suggests that both online summarization practice and answering comprehension questions benefited the learners equally in judging content relevance. This finding can be seen as support for the similar facilitating effects of the online summarization and comprehension exercises. The non-significant difference in grammar and mechanics was attributed to the fact that the researcher/instructor did not provide any instruction focused on grammar rules and mechanics of writing and merely pointed out the common grammatical errors of the learners when discussing the automated feedback with the online summarization group or reviewing the correct answers with the comprehension exercise group. The significantly better performances of the online summarization group in thesis statement and sentence transformation suggest that online summarization practice was particularly helpful for learner improvement in writing a concise thesis statement and paraphrasing the main ideas of the original text. Writing a concise thesis statement and performing satisfactory sentence transformation require a correct understanding of the texts at the macro level and restating the authors' points in one's own words effectively. The findings herein suggest that the online summarization technology provides useful scaffolding for learners to enhance in-depth reading comprehension of texts, and is deemed to be as effective as classroom instruction (e.g., Choy & Lee, 2012; McDonough et al., 2014) in guiding weak learners to develop paraphrasing skills.

In sum, the findings of this study suggest that online summarizing as a learning task is more beneficial to learners than answering short-answer comprehension questions. The opportunities for extensive and repeated practice and the instant automated feedback provided via the online summarization technology contribute to learners' improvements in reading comprehension and summarization.

7. Pedagogical Implications, Limitations of the Study and Suggestions for Future Research

More and more teachers have begun to believe that computer-assisted language learning can supplement classroom instruction when adequately integrated into course syllabi. Online tutoring programs hold great promises for learning outside of classrooms because of the opportunities for repeated practice with instant feedback they provide for learners (Fotos & Browne, 2004). This study extends the findings of previous research on online summarization (e.g., Franzke et al., 2005; Wade-Stein & Kintsch, 2004) by showing that online summarization practice can benefit learners to a larger extent compared to traditional method of answering short-answer comprehension questions. The overall findings of this study suggest that online summarization practice offers a good alternative to comprehension instruction, and can supplement classroom instruction. Further, note that the instructional passages were selected mainly based on the researcher's judgment about the learners' proficiency level. From this perspective, L2 reading teachers should be mindful of learners' English ability when implementing an online program such as WriteToLearn.

This study has a few limitations. First, the participants were EFL learners studying in various fields at a university. They had high cognitive maturity and their English proficiency level was considered average for freshmen of Taiwanese universities in that year. The findings of this study regarding the effects of online summarization can be generalized only to EFL populations with similar characteristics and cannot necessarily be generalized to other EFL populations, such as young EFL learners or college students with low English proficiency. Second, although the online summarization group was not required to revise their first drafts of summaries, many EFL learners in this group did revise their summaries. It was likely that the online summarization group spent more time and effort than the comprehension exercise group in their learning task, and as a result, gained more from training. The findings of this study need to be interpreted with care for the possibly prolonged engagement of the online summarization group.

Further research is needed to substantiate the effectiveness of online summarization practice. An issue worth examining is the extent to which learners with different levels of language proficiency benefit from online summarization practice. Another interesting area to explore would be to investigate how online summarization practice makes learners adopt a more active stance as they read and summarize.

References

- Baleghizadeh, S., & Babapur, M. (2011). The effect of summary writing on reading comprehension and recall of EFL students. *New English Reading Association Journal*, 47 (1), 44-57.
- Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior*, 16 (4), 297-306.
- Bensoussan, M., & Kreindler, I. (1990). Improving advanced reading comprehension in a foreign language: Summaries vs. short-answer questions. *Journal of Research in Reading*, 13 (1), 55-68.
- Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, 25, 313-339.
- Choy, S., & Lee, M. (2012). Effects of teaching paraphrasing skills to students learning summary writing in ESL. *Journal of Teaching and Learning*, 8 (2), 77-89.
- Cordero-Ponce, W. L. (2000). Summarization instruction: Effects on foreign language comprehension and summarization of expository texts. *Reading Research and Instruction*, 39 (4), 329-350.
- Flotz, P. W., Gilliam, S., & Kendall, S. (2000). Supporting content-based feedback in on-line writing evaluation with LSA. *Interactive Learning Environments*, 8 (2), 111-127.
- Fotos, S., & Browne, C. M. (2004). *New perspectives on CALL for second language classrooms*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., & Dooley, S. (2005). Summary Street®: Computer support for comprehension and writing. *Journal of Educational Computing Research*, 33, 53-80.
- Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology*, 26, 3-24.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Jitendra, A., Cole, C., Hoppes, M., & Wilson, B. (1998). Effects of a direct instruction main idea summarization program and self-monitoring on reading comprehension of middle school students with learning disabilities. *Reading and Writing Quarterly*, 14 (4), 379-396.
- Jitendra, A., Hoppes, M., & Xin, Y. (2000). Enhancing main idea comprehension for students with learning problems: The role of a summarization strategy and self-monitoring instruction. *The Journal of Special Education*, 34 (3), 127-139.
- Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, 15 (4), 261-278.
- Kim, S-A. (2001). Characteristics of EFL readers' summary writing: A study with Korean University Students. *Foreign Language Annals*, 34 (60), 569-581.
- Kintsch, W. (1988). The role of knowledge in discourse processing: A construction-integration model. *Psychological Review*, 95, 163-182.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W., Ericsson, K. A., & Patel, V. L. (1999). The role of long-term working memory in text comprehension. *Psychologia*, 42, 186-198.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25 (2-3), 259-284.
- Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). How well can passage meaning be derived without using word order? A comparison of latent semantic analysis and humans. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 412-417). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Landauer, T. K., Lochbaum, K., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory Into Practice*, 48 (1), 44-52.
- Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. New York: Routledge Taylor & Francis Group.

- Malone, L. D., & Mastropieri, M. A. (1992). Reading comprehension instruction: Summarization and self-monitoring training for students with learning disabilities. *Exceptional Children*, 58 (3), 270-279.
- McDonough, K., Crawford, W. J., & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context. *Journal of Second Language Writing*, 24, 20-32.
- Miller, T. (2003). Essay assessment with latent semantic analysis. *Journal of Educational Computing Research*, 29 (4), 495-512.
- Oded, B., & Walters, J. (2001). Deeper processing for better EFL reading comprehension. *System*, 29, 357-370.
- Ono, M. (2011). Japanese and Taiwanese university students' summaries: A comparison of perceptions of summary writing. *Journal of Academic Writing*, 1 (1), 191-205.
- Radmacher, S. A., & Latosi-Sawin, E. (1995). Summary writing: A tool to improve student comprehension and writing in Psychology. *Teaching of Psychology*, 22 (2), 113-115.
- Rinehart, S. D., Stahl, S. A., & Erikson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 21(4), 422-438.
- Rogevich, M., & Perin, D. (2008). Effects on science summarization of a reading comprehension intervention for adolescents with behavior and attention disorders. *Exceptional Children*, 74 (2), 135-154.
- Shokrpour, N., Sadeghi, A., & Seddigh, F. (2013). The effect of summary writing as a critical reading strategy on reading comprehension of Iranian EFL learners. *Journal of Studies in Education*, 3 (2), 127-138.
- Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend texts strategically. In C. Block & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 176-200). New York: Guilford Press.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wade-Stein, D., & Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction*, 22 (3), 333-362.
- Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders*, 30 (4), 275-287.

APPENDIX 1

Scoring Rubric for Summarization Test

Instructions: Each summary will be scored in the following four aspects of summarization. A half point (0.5) may be awarded when appropriate.

	4 points	3points	2 points	1 point
Thesis Statement	<ul style="list-style-type: none"> ⊗ The summary begins with a clear topic sentence that states the main idea of the original text. 	<ul style="list-style-type: none"> ⊗ The summary begins with a topic sentence that states the main idea of the original text. 	<ul style="list-style-type: none"> ⊗ The summary has a topic sentence that touches upon the main idea of the original text. 	<ul style="list-style-type: none"> ⊗ The summary does not state the main idea of the original text.
Content Inclusion & Exclusion	<ul style="list-style-type: none"> ⊗ Major details are stated economically and arranged in a logical order. ⊗ No minor or unimportant details or reflections are added. 	<ul style="list-style-type: none"> ⊗ Major details are stated and arranged in a generally logical order. ⊗ A few minor or unimportant details or reflections are added. 	<ul style="list-style-type: none"> ⊗ Some but not all major details are stated and not necessarily in a logical order. ⊗ Some minor or unimportant details or reflections are added. 	<ul style="list-style-type: none"> ⊗ Few major details are stated and not necessarily in a logical order. ⊗ Many minor or unimportant details or reflections are added.
Sentence Transformation	<ul style="list-style-type: none"> ⊗ The author to a large extent restates the main idea using his/her own words. 	<ul style="list-style-type: none"> ⊗ The author to some extent restates the main idea using his/her own words. 	<ul style="list-style-type: none"> ⊗ The author has some difficulties restating the main idea using his/her own words. 	<ul style="list-style-type: none"> ⊗ The author cannot restate the main idea using his/her own words (copying from the original text).
Grammar & Mechanics	<ul style="list-style-type: none"> ⊗ There are few or no errors in mechanics, usage, grammar or spelling. 	<ul style="list-style-type: none"> ⊗ There are several errors in mechanics, usage, grammar or spelling that do not interfere with meaning. 	<ul style="list-style-type: none"> ⊗ There are some errors in mechanics, usage, grammar or spelling that to some extent interfere with meaning. 	<ul style="list-style-type: none"> ⊗ There are serious errors in mechanics, usage, grammar or spelling that make the summary difficult to understand.