

**SUMMARY STREET[®]: COMPUTER SUPPORT
FOR COMPREHENSION AND WRITING***

MARITA FRANZKE

EILEEN KINTSCH

DONNA CACCAMISE

NINA JOHNSON

University of Colorado

SCOTT DOOLEY

Pearson Knowledge Technologies

ABSTRACT

Having students express their understanding of difficult, new material in their own words is an effective method to deepen their comprehension and learning. Summary Street[®] is a computer tutor that offers a supportive context for students to practice this activity by means of summary writing, guiding them through successive cycles of revising with feedback on the content of their writing. Automatic evaluation of the content of student summaries is enabled by Latent Semantic Analysis (LSA). This article describes an experimental study of the comprehension and writing tutor, in which 8th-grade students practiced summary writing over a 4-week period, either with or without the guidance of the tutor. Students using Summary Street[®] scored significantly higher on an independent comprehension test than the control group for test items that tapped gist level comprehension. Their summaries were also judged to be significantly superior in blind scoring on several measures of writing quality. Students of low-to-moderate achievement levels benefitted most from the tool.

It has long been recognized that comprehension problems can be the consequence of low level reading skills that lack fluency and automaticity (Cornoldi & Oakhill,

*This research was supported by a grant from the Interagency Educational Research Initiative of the National Science Foundation (No. 030.05.0431B).

1996; Perfetti, 1985). Recently, researchers as well as practitioners have realized that there is more to deficient comprehension than weak foundational skills, and that the problem is in fact broader, affecting many students considered to have normal fluent reading skills (Donovan & Pellegrino, 2004; NCES, 2004; Paris & Stahl, 2005).

Failure to read with complete and deep comprehension may interfere with the educational progress of many students throughout their school years and for some even into college. Although adequate decoding skills are the foundation of reading competence, active meaning construction is essential to learning from textual materials. All too often, readers will adopt a passive stance towards instructional reading, especially when reading expository texts about unfamiliar and conceptually difficult topics. Instructional text may inadvertently reinforce passive reading by over-emphasizing ease of readability. Reading and even re-reading texts with short, syntactically simple sentences, largely familiar vocabulary, and concepts that are barely explained and elaborated can create a false sense of understanding. Superficial learning is the result, which leads to poorer than expected performance on tests or in situations that require one to apply what one should have learned. Ultimately, the failure to read with mental engagement becomes a major obstacle to academic progress as students encounter more difficult materials in the later school years.

LEVELS OF COMPREHENSION

Many psychological theories of text comprehension distinguish between superficial versus deep comprehension (e.g., Coté, Goldman, & Saul, 1998; Gernsbacher, 1990; Kintsch, 1998; Trabasso & Suhy, 1993; van den Broek, Young, Ziyang, & Linderholm, 1999; van Dijk & Kintsch, 1983). Comprehension is described not as a unitary process, but rather as a series of interacting levels of processing. In Kintsch's (1998) theory, initial comprehension processes typically involve creating a coherent *textbase* from the words and sentences in the text. This textbase reflects both the local meaning relationships between individual sentences or propositions—the *microstructure*—and the global relationships that pertain between groups of propositions—the *macrostructure*. The textbase is basically what students or participants in an experiment produce when asked to recall the content. No normal text is entirely coherent, that is, with all the semantic relations spelled out. Hence, readers must generate gap-filling, bridging inferences. For example, they must relate pronouns to the proper referent, establish the identity of synonyms, and spell out missing relationships between sentences at the microstructure level; at the macrostructure level they must relate a particular sentence or groups of sentences to a superordinate topic. Depending on a reader's knowledge and verbal skill, such inferences occur quite automatically during reading. Deeper level comprehension involves the construction of a *situation model* of the text meaning—that is, a memory representation that

also includes multiple links to related information in readers' own knowledge base. The situation model thus refers to how the reader understands the situation described in the text in the light of what he or she knows of the topic.

While textbase processing normally proceeds quite automatically, with little conscious awareness, constructing a situation model often requires effortful inferences that resemble problem solving. For example, expert readers will seek ways to connect the text content to what they already know, they paraphrase and construct explanations, and they also monitor their understanding and seek to clarify and fill the gaps. As Kintsch and Kintsch (2005) and others (see report by the Panel on Learning and Instruction, Donovan & Pellegrino, 2004) have argued, getting learners to comprehend what they read at this deep level is a problem that is not for the most part adequately addressed in our schools. Passive reading characterizes many students' encounters with text, which results in inert rather than long-lasting, usable knowledge. The educational software called Summary Street[®] that is the focus of this article is an attempt to foster deep reading strategies; it does so by providing guided support for writing and revising summaries of instructional texts.

Numerous studies, beginning with the ground-breaking research of Ann Brown and her colleagues (Brown, Bransford, Ferrara, & Campione, 1983; Brown & Day, 1983) have demonstrated the effectiveness of summary writing as one way to enhance comprehension and learning. With respect to the comprehension theory briefly described above, a summary is an overt expression of the macrostructure a reader has constructed during reading. However, the process of composing a written summary requires a good deal more effort than reading or even re-reading a text. Summarizing is a form of meaning construction: one must decide which content to include, focusing on the most important, main ideas; one must restate the content in a more concise form, using one's own words, balancing the need to be complete with the need to be brief. In so doing, a summarizer works with the meaning of the text at a deep level, comparing the text meaning with how it has been conveyed in the summary, gradually making the new knowledge his or her own. Thus, summary writing can be used instructionally as a means to promote understanding at the level of both the textbase and the situation model.

While teachers acknowledge that having students articulate their understanding by writing summaries is an excellent way to encourage active, engaged reading and deep understanding, the task of reviewing and giving individualized feedback is often seen as an overwhelming burden to them. Hence, students rarely get enough opportunities to practice writing and revising their summaries.

SUMMARY STREET[®] AND LSA

The comprehension tutor called Summary Street[®] seeks to address this problem by providing a supportive context in which students engage deeply with text meaning by expressing their understanding in written summaries. The tutor

provides nearly instantaneous evaluation of the content adequacy of students' summaries, regardless of the exact words used. The analysis of text meaning is made possible by the statistical method called Latent Semantic Analysis (LSA) developed by Landauer and Dumais (Landauer, 2002; Landauer & Dumais, 1997). By analyzing how words are used in a very large number of texts, LSA automatically constructs a representation of meaning that mimics the way human knowledge is structured. The knowledge representation constructed by LSA is, of course, incomplete in many ways, since its knowledge is acquired only from written texts. However, since texts are mainly about human experiences, perceptions, and conceptualizations, given a sufficiently rich textual input, LSA can judge the meaning similarity of two texts as reliably as two human judges (Landauer, Laham, & Foltz, 2003). Introductions to the LSA method can be found in Landauer (2002) and Landauer and Psotha (2000) for readers interested in learning more about how it works. See also Miller's (2003) review of LSA and LSA-based educational tools, which compares this method to other statistical methods for evaluating text similarity.

The potential of LSA for supporting educational goals was quickly grasped and has been incorporated into a growing body of tools, including, for example, grading college student essays (IEA), cross-language document retrieval, assigning books or training to learners appropriate to their level of skill and knowledge. An overview of these and other applications is available in Landauer, McNamara, Dennis, and Kintsch, (in press) and at www.PearsonKT.com.

In our educational application called Summary Street[®], LSA is used to compare the similarity in meaning between a student's summary and the source text from which it was derived. When a student submits a summary to Summary Street[®] for evaluation, it is transmitted via the Internet for analysis at the LSA servers. Feedback is returned almost instantly in the form of an engaging graphic interface, which lets the user know to what extent the summary adequately covers the main ideas and what topics need more work. The student uses the feedback to improve his/her summary through successive cycles of revising and feedback requests until it satisfies the criteria for content coverage and appropriate length. Figure 1 shows a screenshot of the feedback page: the horizontal bars correspond to the section headings in the to be summarized text; the vertical bar on the right provides a length guideline. For each topic section, LSA computes a cosine (similar to a correlation coefficient) as a measure of similarity between the information in the summary about that topic and the source text. Note that LSA does not rely on the presence of keywords, word overlap, or synonyms in the two texts, but rather reflects their semantic relatedness, regardless of the actual words used (Landauer, 2002; Landauer & Psotha, 2000).

An algorithm is used to determine the threshold for content coverage for each main topic or section of the text, which varies in each case; however, students are unaware of these numbers and see improvement in content coverage in terms of how closely the horizontal bar approaches the vertical black threshold line for each

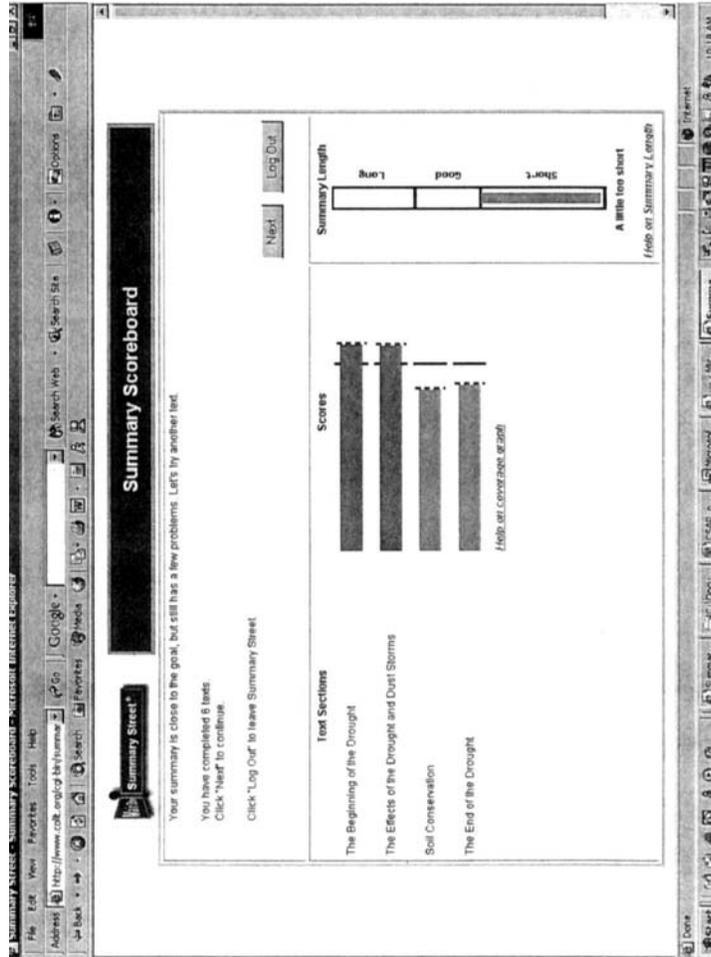


Figure 1. Summary Street window, showing feedback screen for a summary of "The Dust Bowl." The lower two bars are yellow, indicating incomplete section coverage; the upper two bars are green, indicating adequate section coverage. The vertical bar is a length indicator, which is green if summary length is in the appropriate range and yellow if it is too long or too short.

section. Color provides an additional cue to the adequacy of content coverage: a yellow bar indicates inadequate coverage of a section, while green indicates that the threshold has been passed or surpassed for that section. The summary length is typically set by the teacher to be between 10% to 25% of the original text; summaries that are too short or too long are indicated by the orange vertical bar, which likewise turns green when summary length is in the acceptable range.

In addition to content coverage, students can request a spell check as well as checks for possible problem sentences—sentences that are overly redundant with other sentences in the summary or that are not very relevant to the topic. Plagiarized sentences are also flagged. A detailed description of how LSA functions in Summary Street[®] and issues regarding this approach to meaning evaluation is available in Wade-Stein and Kintsch (2004). Kintsch et al. (2000) provide a history of how the tool was developed. It is important to emphasize that the feedback provided by Summary Street[®] focuses on the content of the writing: whether all the main topics are adequately covered in a summary and whether the information is concisely stated, relevant, and not copied from the source text. Summary Street[®] does not directly evaluate other aspects of writing quality, such as organization, writing style (e.g., voice), and punctuation and sentence structure. When supported by the system, students can do a lot to improve the quality of their summaries on their own, simply by attending to content matters before handing it in for a teacher's final evaluation.

Summary Street[®] is currently in use in some 70 5th- through-11th-grade classrooms in the Denver metro area, where large-scale evaluation is underway. The beneficial effects of using Summary Street[®] on the quality of students' written summaries has been shown in classroom trials by Wade-Stein and Kintsch (2004), who found that 6th-grade students wrote better summaries when guided by the content-based feedback the tool provided than when using a word processor that only provided information on the length and spelling. Moreover, the authors found that the students spent more than twice as much time working on their summaries when using Summary Street[®]. Summary Street[®] helped the students give equal attention to all the main topics in the source text and to include this information in their summaries. Without this guidance, students' summaries typically focused more on the early topics and neglected those that occurred later in the source text.

In the present study, we seek to replicate these results using different texts and a somewhat older student population and to see whether the beneficial effects of Summary Street[®] use may accrue over a longer period of practicing with the tool. In addition, the study provides a more detailed look at what aspects of summary writing are benefitted by the feedback from Summary Street[®]. The question of individual differences is also explored, as it is possible that students of differing ability levels will respond in different ways to the feedback.

Finally, we seek to provide evidence for the hypothesis that sustained, guided practice in summary writing improves the way students read and comprehend

difficult textual materials, as measured by an independent test of comprehension: does practice in summarizing, guided by feedback on the content of one's writing, enhance comprehension processes beyond what is gained from simply practicing summary writing without feedback? Supporting the process in this way over a period of several weeks by means of automatically delivered feedback on content coverage may make students read more attentively, with better awareness of the important information in a text when answering questions on a comprehension test.

As an independent test of the students' comprehension, we selected comprehension questions taken from the Colorado Student Assessment Program (CSAP) that were no longer in use and were publically available. The CSAP is a state mandated standards test administered to public school students in grades 3 to 12. Standardized tests, such as the CSAP, are now used in all states by the educational authorities to achieve accountability at the district and school level. In many states, including Colorado, such tests are constructed on the basis of specified standards of educational competency in reading comprehension, writing, mathematics, and science. Although these are standardized tests, they include many open-ended as well as multiple-choice test items. Schools and teachers are familiar with the format of the tests, and an improvement on reading comprehension items from the CSAP test would provide convincing evidence of the benefits of Summary Street®. As the current study was conducted within the framework of a project that had dissemination and evaluation of the software in the Colorado Public School system as its main goal, the CSAP seemed like a suitable choice to measure comprehension performance in a meaningful way.

Standard tests of reading comprehension, such as the CSAP, are composed of different kinds of questions, only some of which are related to the skills that Summary Street® supports. Hence, transfer was expected for summary questions because they directly target gist-level comprehension of the text content. Transfer on inference items might also be expected, which would suggest that deeper processing of the text through guided summary writing would help students apply their knowledge better. On the other hand, questions that target specific facts, literary interpretation, analysis of writing style, and vocabulary knowledge are less likely to be influenced by Summary Street® feedback with its focus on text macrostructure. Thus, we are comparing two broad categories of pre- and post-test items: those for which we expect to find an effect of the content feedback versus those for which no effect is expected.

METHOD

Participants

Eighth-grade students from four language arts classes were recruited to participate in the study. All classes were taught by the same teacher (MAS) at Century Middle School in Thornton, Colorado. Century Middle School is located in a

suburban area of the Denver Metropolitan area, with a primarily white lower-middle-class to middle-class population. The proportion of minority students (Hispanic, American-Indian, Asian, and African-American is roughly 24 percent, as measured by the Colorado Department of Education 2002 student census. Across all four classes, there were 121 students who participated. However, the analyses are based on responses of 111 students who missed no more than three of the eight summary practice sessions. Students in these four classes ranged in age from 13.5 to 14.5 years. Of the participants with usable data, 52 were in the Summary Street[®] condition (24 females and 28 males) and 59 participants were in the control condition (38 females and 21 males).

Three of these classes were regular language arts classes; the fourth class was designated for high-performing students. Students' overall CSAP scores, collected in the spring of 2002, indicated that their achievement level on these measures was quite similar across all four classes. CSAP scores of students in the higher-performing class tended to be at the higher end of the distribution, but not beyond the range of students in the other classes. CSAP scores of student distribution, but not beyond the range of students in the other classes. Thus, assignment to experimental and control condition was made within classes, with roughly half of the children in each class randomly assigned to either condition. The mean CSAP scores for the experimental group and control group were 650.92 and 641.98, respectively ($t_{(97)} = .33, p > .05$); standard deviations were 50.02 and 43.09, respectively, indicating that the two groups were comparable in reading achievement levels.

Materials

The intervention took place during the unit on summary writing, as part of the regular 8th grade language arts curriculum that focuses on learning about and practicing different writing genres (e.g., poetry, various types of narratives, creative writing, informative writing).

CSAP Test Materials

We constructed a comprehension test similar to the one the students would be tested with later that spring, one consisting of both open-ended as well as multiple-choice test items. For this purpose we selected retired, published comprehension items from the Colorado Student Assessment Program (CSAP) to test comprehension performance of the students prior to and following the summary practice sessions. Because of the pre- and post-test design of our study, two comparable sets of comprehension items were constructed in order to avoid practice effects: each student saw a given item only once. Selection of the test items was based on the following criteria:

1. All test items were in the open domain and accompanied with a scoring guide and examples published by the Colorado Department of Education

standards; year of administration and grade level of the items precluded a student having previously seen any of the items.

2. The two versions of the test (called CSAP-A and CSAP-B) were comparable in terms of CSAP grade level, with test items spanning grades 7 to 10.
3. They contained a similar distribution of text genres (expository, biography, fiction, and poetry).
4. The two tests were comparable in terms of number of test items and number of multiple choice and constructed response questions.

To analyze students' comprehension performance, the test items of the two test versions were first categorized into the following groups: summary questions, inference questions, vocabulary questions, fact-finding questions and a few questions, labeled "other," that did not fit these categories. Effects of Summary Street[®] use were anticipated for summary and inference questions, while no effects were expected for any of the other item types.

As much as possible, the two test versions were equated in terms of the numbers of each question type, but this was not possible in all cases, given that these are actual test items, not constructed especially for an experimental trial. Table 1 shows examples for each of these item types.

Summary Practice Materials

To practice summary writing, the students in this study read and summarized a selection of 19 short-to-medium length texts. These texts were assembled into a practice booklet that was given to both the experimental and the control participants. Students in both conditions worked on the same texts in the same order. The booklet also included a brief introduction to summary writing and, for the experimental group who would be using Summary Street[®], a brief overview of how to use the tool (interested readers can view the guidelines and try out the system as a guest user on the Summary Street[®] Website: www.colit.org).

The 19 practice texts were selected from standardized comprehension tests that had been published on the Internet without copyright restrictions. In a few cases, permission to use copyrighted texts in our study was obtained from the publisher. Texts were selected to be similar in style to those used in the CSAP tests and to be generally appropriate in subject matter and difficulty for the students in this study. A range of topics and genre types were represented: expository texts on scientific or social issue topics, stories, and biography. The practice texts ranged in difficulty from 7th to 12th grade and were between one page to several pages long (range = 625 to 1682 words). In addition to the grade level designations of the test publishers, the texts were rated by the researchers for conceptual difficulty of the subject matter. Section headings were provided for texts that lacked a breakdown into sections to make them more suitable for use with Summary Street[®]. Using these criteria, we ordered the texts such that the students would work generally from shorter, easier texts with simpler subject matter to longer

Table 1. Categories of CSAP Comprehension Items

Item type	Example
Summary	Write a summary of Sojourner Truth's Speech.
Inference	<p>Why did Drew's plasma research become especially important during World War II?</p> <p><input type="checkbox"/> Whole blood could now be given to anyone.</p> <p><input type="checkbox"/> Many people were injured and needed blood.</p> <p><input type="checkbox"/> Shipments of blood were halted by enemy attack.</p> <p><input type="checkbox"/> Military doctors were required to do more medical research.</p>
Fact Finding	<p>List four important facts about Sojourner Truth that you read in the article.</p> <p>1. _____</p> <p>2. _____</p> <p>3. _____</p> <p>4. _____</p>
Vocabulary	<p>Read these two sentences from the article:</p> <p>"Yes!" cried someone near the front. "True!" voices reverberated across the church.</p> <p>What does the word reverberated mean?</p> <p><input type="checkbox"/> echoed</p> <p><input type="checkbox"/> heckled</p> <p><input type="checkbox"/> laughed</p> <p><input type="checkbox"/> whispered</p>
Other: lit. analysis	<p>Below are two examples of figurative language that the author uses to describe Sojourner Truth. Explain how each comparison helps the reader to understand Sojourner Truth.</p> <p><input type="checkbox"/> "stately as a queen"</p> <p><input type="checkbox"/> "her voice rising like a roll of thunder"</p>

texts with more sections that described more sophisticated content. We also ordered the texts such that the students would be exposed to a mixture of text styles by interspersing expository texts, biographical material, and narratives as evenly as possible. Table 2 shows the first six texts in the series (T1-T6), with associated text metrics. These are the texts that a majority of students summarized. Four expository texts were selected from this set for scoring and analysis: T1 "Springsteen"; T3 "Vaccination"; T4 "A Girl with Stars in Her Eyes"; and T6 "Dust Bowl" (highlighted in Table 2). These form a sequence of texts that are

Table 2. Characteristics of the Six Texts Completed by the Majority of Students

Order	Title	Type	Word count	No. of sections	Grade level	Conceptual difficulty
T1	Springsteen Concert Debated	expository	696	1	8	easy
T2	The Path Through the Cemetery	narrative	626	1	8	easy
T3	Vaccination	expository	625	2	8	moderate
T4	A Girl with Stars in Her Eyes	biography	1022	4	8	moderate
T5	Cady's Life	narrative	868	1	8	moderate
T6	Dust Bowl	expository	954	4	12	moderate

Note: Estimated grade level provided by the publisher; conceptual difficulty rated by researchers. Texts shown in **bold** are those used in the analysis.

similar in genre and make similar though increasing demands on the writer. Summaries of the stories (T2 and T5) were excluded from this set because they require different processing strategies. Although T4 "A Girl with Stars in Her Eyes," a biography, has an overall narrative structure, it includes long descriptive passages about astronomy that make it comparable to the other expository texts.

Procedure

Pre-Test and Post-Test Administration

Comprehension performance was measured before and after the intervention using the two CSAP tests (CSAP-A and CSAP-B) described above. Half of the students answered CSAP-A during the pre-test, then answered CSAP-B during the post-test, and the other half of the students answered the item sets in the reverse order. The order of presentation of CSAP-A and CSAP-B across experimental groups, classes, gender, and ability level was random.

The CSAP pre-test was administered in the four classrooms the week before the winter break during two 45-minute sessions. The post-test was administered in the week following the last summary practice session, also during two 45-minute sessions. For both the pre-test and a post-test, students were provided with a booklet of items and instructed to work through it at their own pace.

Summary Practice Sessions

The intervention took place in four language-arts classes in twice-a-week sessions during a 4-week period, in lieu of the teacher's regularly scheduled instruction and practice in summary writing. It is important to note that students in both conditions were writing summaries of the same texts in the same order during this time, using the same general instructions, but only the Summary Street[®] students received feedback on their summaries. Students in the control group wrote their summaries using a word processing program. Summary writing sessions for each class were scheduled in a computer laboratory in the school library. Students in both conditions used the same lab, but were grouped in different areas of the lab to avoid too much cross-stalk among students in differing conditions. They were all provided with hard copies of the guidelines on summary writing mentioned above. All students were asked to work at their own pace, and to work on the texts in a specified order. They were informed about the research nature of this project, but were told that the summaries they produced would be used for class credit, no matter whether they used Summary Street[®] or a word processor to produce their summaries. They were informed that the number of summaries submitted would not be used to judge their performance, but that the quality of the summaries would be important. All students were asked to submit their six best summaries for grading by the teacher at the end of the intervention.

Students in the Summary Street[®] condition received some one-on-one instruction on the use of the program during the first two sessions, depending on how much help was needed by individual students. Most questions concerned how to submit and save summaries for future access. The Summary Street[®] software was modified slightly to meet the needs of the current use situation. To insure that students in the Summary Street[®] condition would not spend an inordinate amount of time on each text, we introduced more fluid rather than absolute thresholds for content coverage, which would allow students to pass the overall criterion after a given number of attempts if they had gotten close to the threshold, but had not actually met it. Thus, for example, a student would pass the criteria for a given text if the following conditions were met: a) the content score for every section was within 5% of the threshold; b) the student had already submitted the summary 10 times for content feedback; and c) the summary was within the specified length requirements. Students in the experimental condition were instructed to work on the summaries at their own pace, using the feedback provided. After passing the Summary Street[®] thresholds, as described above, they were instructed to work on the next text in the sequence indicated on the computer screen.

Students in the control condition were told to use a word processing program to write the summaries to the best of their ability and at their own pace. They could access the functions provided by the word processor to check spelling and

length. Their summaries were stored in a public folder, and these students also received help with storing and retrieving their summaries during the first few sessions. After receiving these instructions students worked independently.

Post-trial Interviews

On the last day of the trial, following the comprehension post-test, we conducted interviews with 18 students who had volunteered for this extra task; 10 interviews were with students who had been in the experimental and 8 with students from the control group. The interviews were semi-structured and focused on students' perceptions and observations regarding summary writing (both conditions) and the use of Summary Street® in particular (the experimental group).

RESULTS

Summary Completion Rates

On average, students in the experimental condition summarized 6.55 texts, with a standard deviation of 1.72, whereas students in the control condition summarized an average of 5.75 texts with a standard deviation of 2.40. According to a *t*-test, the means from these two populations did not differ significantly ($t_{44} = -1.55, p = .12$). This result indicates that students in both conditions made fair attempts at the task and proceeded through the texts at approximately the same pace of a little less than one text per session. Summary completion rate remained stable across the four weeks as the texts increased in length and complexity.

Comprehension Pre- and Post-Test

Five raters scored the CSAP pre- and post-test results from 111 students using the official published CSAP grading rubric and examples. Inter-rater reliabilities were calculated for all pairs of graders. The average of inter-rater correlation was .86.

Analysis 1

Scores from the two versions of the test (CSAP-A and CSAP-B) were used to examine differences between experimental and control groups for overall performance and for each item type (summary, inference, fact finding, vocabulary, and other). However, since versions A and B of the test did not contain the same number of items of each type, a scaling procedure was employed to account for differences between test forms. Scores were scaled using Angoff's method (1971), which equates the difficulty of the two versions by using the pre-test means as an indicator of difficulty differences between formats. Comparison of mean scores between experimental and control groups employed

Factorial Analyses of Covariance (ANCOVA) with the pre-test used as a statistical covariant. Class membership was also used as an independent variable. A significance level of $p < .05$ was adopted throughout.

Table 3 presents the mean post-treatment scores for overall performance and broken down by particular item type. Note that the number of points differed in each category, as shown in the table. Because of the statistical procedure used (ANCOVA), only cases with matched pre-/post-test scores are included in the analysis, which leads to small differences in the degrees of freedom for different scores.

With respect to the main effect of condition, we found no significant improvement in experimental and control groups on overall post-test scores, nor on scores for inference, fact finding, vocabulary, and other questions. However, scores between the two conditions did differ on one item type, namely those requiring a summary response. Using Summary Street® helped the experimental subjects to score significantly better on these test items: the increase from 30% to 38% correct represents a gain of 25% from pre- to post-test, corresponding to an effect size of .42.

With respect to the main effect of class, we found that class membership explained a significant amount of variance for overall scores, fact-finding items, and inference items. This result was not surprising given that one of the four classes was a high-performing class, as discussed previously. As could be expected, that class outperformed the others on these measures. However, there was no significant class-by-condition interaction on any of the measures, showing

Table 3. Post Intervention Comprehension Test Scores per Item Type

Categories	No. of points ^a	Summary Street ^b	Control group ^b	ANCOVA
Overall	37	22.90	22.05	$F_{(1,83)} = 0.34$
Summary	7	2.64	2.12	$F_{(1,83)} = 4.05^*$
Inference	13	7.95	7.72	$F_{(1,85)} = 0.24$
Fact finding	9	4.90	4.45	$F_{(1,87)} = 1.39$
Vocabulary	4	2.15	2.22	$F_{(1,90)} = 0.53$
Other	4	3.08	2.73	$F_{(1,91)} = 0.81$

^aNumber of possible points for each item category.

^bMean number of points correct for Summary Street and control groups.

* $p < .05$

that whatever differences due to the use of Summary Street[®] existed, they were not modified by class membership. In other words, students performed better on the summarization items when they used Summary Street[®], no matter which class they belonged to.

Analysis 2

A second analysis of the comprehension test data, which did not require scaling for the differences between the two versions of the test, was performed by computing the change from pre- to post-test for each item. Overall scores on the test items revealed no significant difference between pre- and post-intervention performance: neither group's *overall* performance changed as a result of the intervention. However, the interaction of group and item type revealed a significant difference for items requiring summary responses. The scores of Summary Street[®] users improved on these questions, while the scores of the students in the control group did not. The difference between the two groups is statistically significant as measured both by a sign test ($p < .05$) and a paired t -test ($t_6 = 2.62$, $p < .05$). There was no significant difference in pre- to post-test change scores for any of the other item categories.

In sum, both types of analyses of the comprehension test results show an identical pattern of results: there was no difference between experimental conditions in overall comprehension scores, nor for inference, fact-finding, vocabulary, or other test items, but subjects using Summary Street[®] performed significantly better on the summary items than subjects who had no access to Summary Street[®]. Thus, training with Summary Street[®] enabled subjects to transfer their skill in summary writing to an independent testing situation: they performed significantly better than control group students, who practiced summarizing an equivalent amount of time, on items of the CSAP comprehension test that required summary responses.

Summary Quality

Scoring

Feedback from Summary Street[®] should also affect the quality of students' summaries. Hence, the quality of students' summaries was analyzed for the four expository texts—T1 "Springsteen," T3 "Vaccination," T4 "A Girl with Stars in Her Eyes," and T6 "Dust Bowl."

Table 2 provides an overview of the text characteristics for these four texts and shows that they form a sequence in terms of difficulty level. We had purposefully started the series of practice texts with two short texts that consisted of a single section each, followed by longer texts with more sections and conceptually more difficult material. For the description of the results that follows, it is important to keep this fact in mind. As the texts become more difficult, no overall

improvement with practice should necessarily be expected. If scores remain at the same level despite increasingly complex material, it actually implies improvement—even though this may not be evident in students' raw scores.

To gain a detailed picture of the effects of Summary Street® use on summary quality, our analysis examined differences related to both group and ability level on several measures of writing quality. Two teachers (not otherwise involved in the study) each scored half of the summaries used in this analysis. Each set contained almost equal numbers of summaries from the control and experimental groups and for each of the source texts. Any identifying information was removed from the copies of the summaries so that experimental and control summaries could not be distinguished by font or any other external cues. Thus, the teachers scored the summaries blindly.

The teachers were provided with grading categories and some general instructions on the criteria to use in grading the summaries, but did not communicate with each other or jointly score a subset of the summaries. Additionally, 32 individual summaries were double-scored to compute inter-rater reliabilities for the two graders. The grading rubric was similar to the one used in Wade-Stein and Kintsch (2004). Teachers were asked first to evaluate the summaries on each of the following criteria, using a scale of 1 to 5 (with 5 being the best grade possible):

- Holistic Quality
- Organization
- Mechanics (sentence structure, punctuation and spelling)
- Minimal Use of Detail
- Writing Style

Second, for each section of a given text, the scorers judged whether the text content had been covered adequately in the summary on a scale of 0 to 2 points (with 2 being complete and appropriate coverage). Finally, they also provided a score for plagiarism on a scale of 0 to 2 (with 0 being no evidence of plagiarism, and 2 being almost entirely or entirely copied from the text). The computed inter-rater reliability for the holistic grade was .50, which seems low, but is not unusual for holistic scoring between raters who have not trained together (Shanteau, 2001).

Student ability levels were derived from the state-mandated CSAP test scores collected by the school district at the end of the previous school year: students scoring in the lower, middle, and top third were classified as low, medium, or high comprehension level, respectively, based on their scores. Analyses of variance with orthogonal contrast codes were computed for all scores provided, treating the four texts as within subject, and condition (control, Summary Street®) and ability level (low, medium, high) as between subject variables. A significance level of $p < .05$ was adopted for these analyses.

Effects of Condition on Summary Quality

The results of the analyses of summary quality are shown in Table 4 and depicted graphically in Figures 2 and 3. For ease of reference, the results of the individual measures of scoring quality are divided between two figures: those for which positive benefits of Summary Street® feedback were expected, and those where no effects were anticipated. As seen in Figure 2, our predictions were confirmed for holistic quality, content, organization, and minimal use of detail, but surprisingly, the benefits also extended to the style score, shown in Figure 3.

Table 4 shows, in more specific detail, a strong main effect of condition when all four texts are averaged together, such that the students who had produced the summaries with the help of Summary Street® wrote summaries that were judged to be better in overall quality and more complete in their coverage of the text content. A significant interaction of text and condition was found in

Table 4. Means, Standard Deviations, and *F* Values for Summary Quality Scores

Measures	Main effects			Interaction of condition and text				
	Summary Street		<i>F</i>	Summary Street		Control		<i>F</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
Holistic quality	3.32 (1.06)	2.95 (1.30)	*	3.18 (1.05)	3.48 (1.04)	3.14 (1.27)	2.68 (1.30)	**
Content	1.51 (.55)	1.28 (.60)	**	1.48 (.66)	1.55 (.42)	1.38 (.62)	1.14 (.54)	**
Organization	3.58 (1.09)	3.39 (1.33)		3.38 (1.17)	3.79 (.96)	3.46 (1.37)	3.30 (1.28)	**
Mechanics	3.48 (.92)	3.45 (1.18)		3.44 (1.00)	3.53 (.84)	3.46 (1.18)	3.45 (1.19)	
Detail	3.65 (1.04)	3.29 (1.16)		3.40 (1.08)	3.91 (.93)	3.21 (1.22)	3.40 (1.08)	**
Style	3.29 (1.12)	2.98 (1.29)		3.12 (1.09)	3.47 (1.14)	3.06 (1.32)	2.87 (1.26)	**
Plagiarism	.52 (.78)	.46 (.74)		.41 (.74)	.63 (.81)	.44 (.71)	.49 (.78)	

Note: High score for amount of detail denotes few details.
p* < .05. *p* < .01.

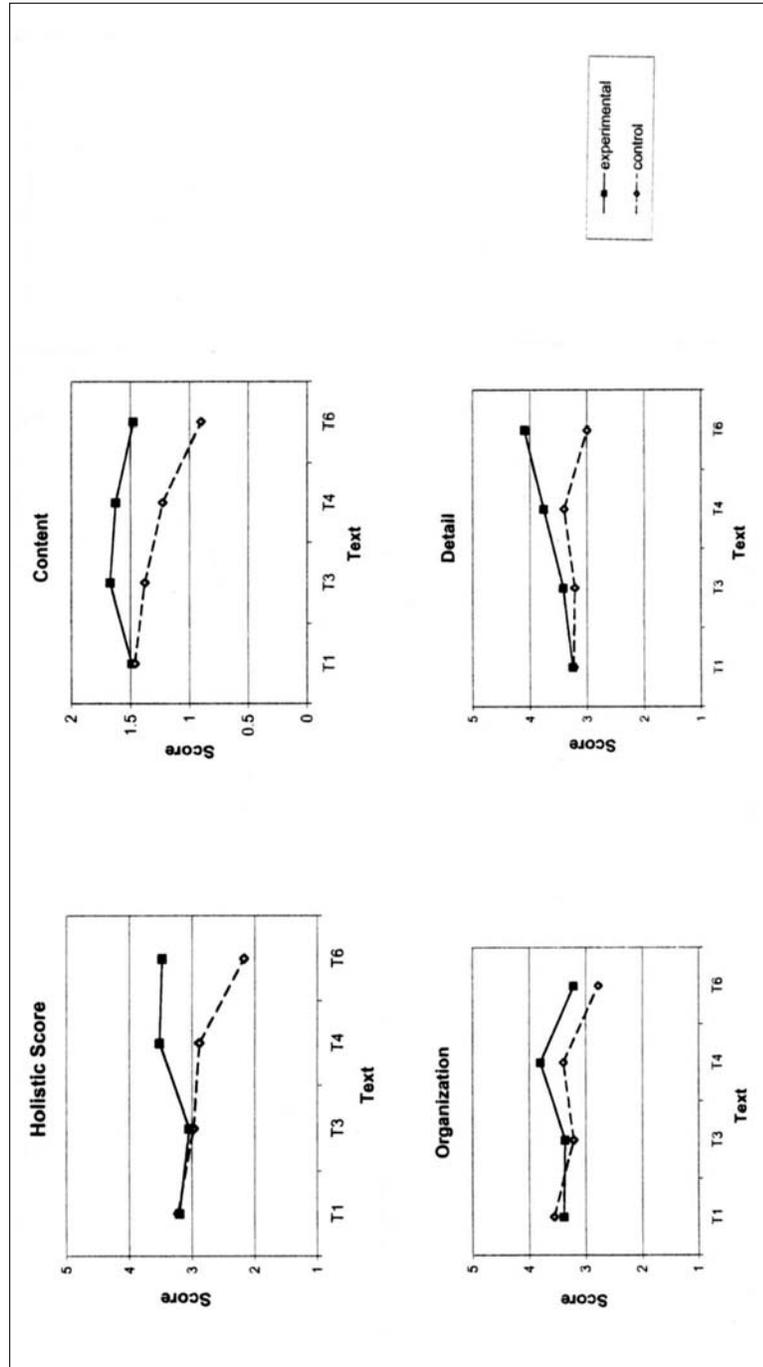


Figure 2. Measures of summary quality—holistic, content, organization, and detail scores—as a function of condition and text.

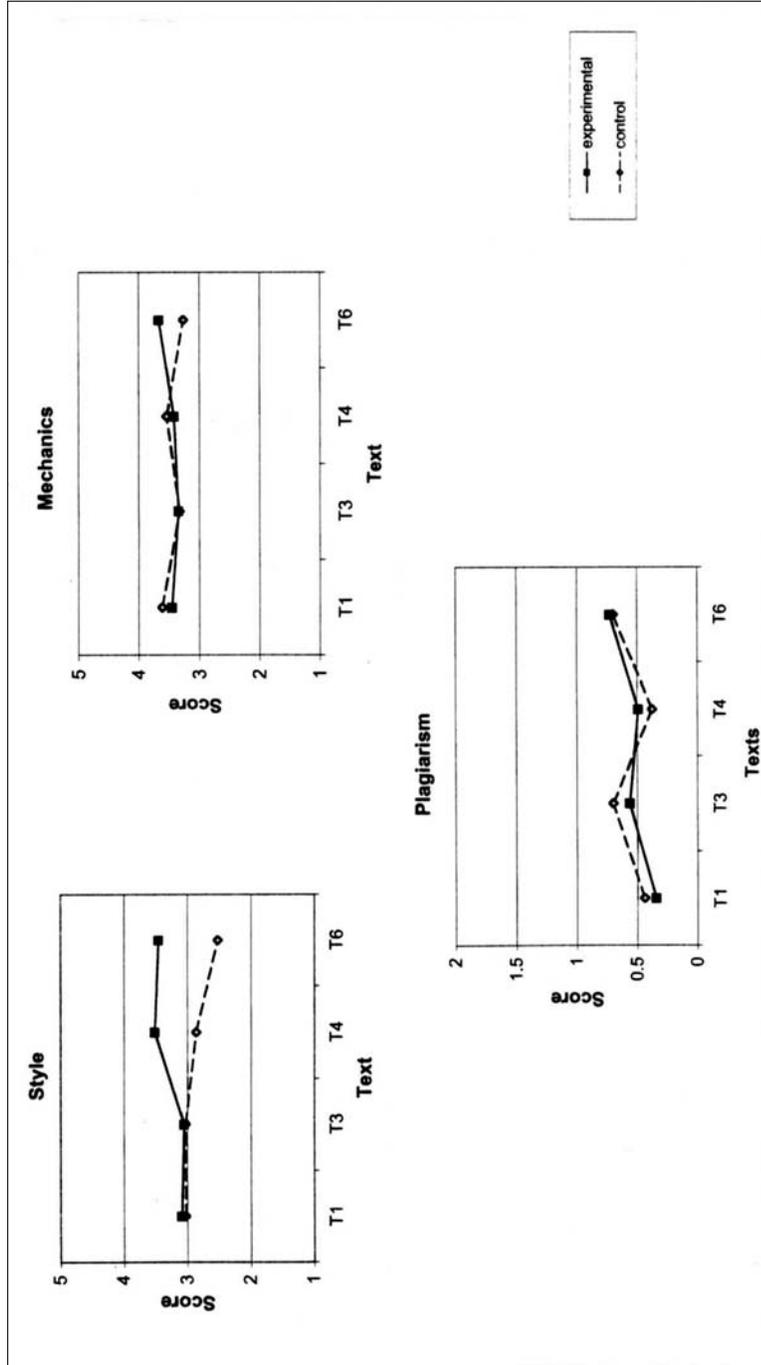


Figure 3. Measures of summary quality—style, mechanics, and plagiarism scores—as a function of condition and text.

contrasting the earlier texts in the series averaged together (T1 and T3) with the average of the later texts (T4 and T6). This finding indicates that across time summaries composed with Summary Street® improved significantly compared to the control group summaries, not only in terms of quality and content but also in terms of their organization, the low amount of detail, and even their stylistic quality. The content-level feedback apparently helped students write better summaries across the board, as shown by several important indicators of writing quality, but more so after some practice and even as the source texts became longer and more complex. Summary Street® users were no better than control students in their attention to the mechanics of writing—spelling, punctuation, and sentence structure—as reflected in the non-significant difference in mechanics scores. No differences were expected on this measure. Spelling is a major problem for this age group; however, both groups had equivalent access to spelling help. Summary Street® also does not encourage excessive sentence lifting from the source text, as both groups had similarly low scores on plagiarism. These results are shown in Figure 3 and Table 4. Additional orthogonal contrasts comparing T1 with the average of T3, T4, and T6 and also of T1 with T3 were consistent with the pattern described above, showing significant interactions of condition with text for scores measuring holistic quality, content, organization, minimal use of detail, and style.

Differences in Ability Level

The effect of ability level on summary writing was reflected in significant or close to significant differences for six out of seven measures of summary quality—all except the plagiarism score. These results are presented in Table 5. Considering the average scores across all four texts, we see that students who scored in the upper third on the CSAP test outperformed the rest of the students on all of the summary quality measures, as shown by the contrast of high versus the averaged scores of medium- and low-performing students. Overall, the high ability students were more likely to include relevant content; their summaries were better organized; they included few irrelevant details; they received higher style scores; and they received better scores for mechanics and spelling, leading to better scores for overall quality. Similarly, a second contrast code, comparing the performance of medium- to low-ability students, showed significantly higher scores for the former group for holistic quality, content coverage, and organization. Thus, the general ability level of the students seems to be reflected in the quality of their summaries.

While the best students outperformed the medium- and low-performing students when results are averaged across condition and text (see Table 6), a pattern of significant interactions of ability level, condition, and text occurred. This finding suggests that Summary Street® was especially beneficial to medium- and low-performing students, although it helped everyone's performance to some degree.

Table 5. Means, Standard Deviations, and *F* Values for Summary Quality Scores by Ability Level

Measures	Contrast Code I			Contrast Code II		
	High <i>M (SD)</i>	Medium + Low <i>M (SD)</i>	<i>F</i>	Medium <i>M (SD)</i>	Low <i>M (SD)</i>	<i>F</i>
Holistic quality	3.62 (1.00)	2.93 (1.20)	+	3.15 (1.12)	2.54 (1.23)	*
Content	1.60 (.43)	1.32 (.63)	*	1.41 (.50)	1.15 (.78)	**
Organization	3.96 (.96)	3.27 (1.24)	+	3.51 (1.17)	2.85 (1.27)	*
Mechanics	3.93 (.73)	3.24 (1.09)	**	3.40 (1.05)	2.97 (1.12)	
Detail	3.86 (.96)	3.30 (1.12)	*	3.42 (1.07)	3.10 (1.18)	
Style	3.55 (1.08)	2.95 (1.22)	*	3.17 (1.19)	2.58 (1.18)	
Plagiarism	.41 (.77)	.51 (.76)		.47 (.72)	.60 (.82)	

* $p < .05$. ** $p < .01$. + $p < .06$.

Specifically, as seen in Figure 4, the performance gain across 4 weeks of summarizing practice for students who used Summary Street[®] was larger for medium- and low-ability students than for high-performing students, for the holistic quality, content coverage, and organization scores. Content and organization are the areas where Summary Street[®] provides most direct feedback, and the lower performing group of students could apparently use this feedback to significantly improve the quality of their summaries. When the same set of analyses was run comparing the summary scores of low- and medium-performing students, none of these interactions was significant. Thus, in Figure 4 the scores for the low- and medium-ability groups are averaged. Interactions for minimal detail, style, mechanics, and plagiarism scores were not significant in the analyses.

In sum, the analyses of the summary quality data show that Summary Street[®] helped all students write better summaries than students in the unsupported control condition, according to several indicators of quality. These effects became more pronounced with more practice, even as the texts became longer and more complex: Summary Street[®] feedback served to remind students to include relevant

Table 6. Means, Standard Deviations, and *F* Values for Summary Quality Scores by Condition, Text, and Ability Level

Measures	Summary Street				Control				<i>F</i>
	T1 + T3		T4 + T6		T1 + T3		T4 + T6		
	High	Medium + Low							
	<i>M</i> (<i>SD</i>)								
Holistic quality	3.26 (.62)	2.94 (1.28)	3.62 (.97)	3.41 (1.08)	3.86 (.89)	2.76 (1.28)	3.25 (1.24)	2.39 (1.26)	*
Content	1.63 (.43)	1.41 (.73)	1.60 (.39)	1.52 (.43)	1.70 (.43)	1.21 (.64)	1.41 (.47)	1.01 (.53)	**
Organization	3.96 (1.07)	3.12 (1.14)	3.76 (.94)	3.81 (.97)	4.32 (.84)	3.00 (1.38)	3.75 (.93)	3.06 (1.39)	*

* $p < .05$. ** $p < .01$.

content from all sections of the texts. As a result, their summaries tended to be better organized and contained relatively little low-level detail. Multiple cycles of revising apparently helped them improve their summaries from a stylistic standpoint as well. Furthermore, even though higher ability students wrote better summaries in general, when scores were averaged across text and condition, we found that Summary Street[®] was especially effective for medium- and low-ability students, allowing them, after some practice, to approach the performance level of the high ability students.

Interview Results

On the last day of the trial, following the comprehension post-test, interviews were conducted individually with 18 students who volunteered for this task. Ten students were from the Summary Street[®] group and eight from the control group. Most of the interview questions focused on usability issues that will not be covered here. However, the final part of the interview, in which each student was asked to state what he or she thought was most important in producing a good summary, provides some additional insight into how Summary Street[®] benefitted the students who had used it. In responding to this open-ended question, students who had used Summary Street[®] mentioned at least two important summarization strategies on average, whereas the control group students typically mentioned only one. Moreover, Summary Street[®] users mentioned strategies bearing directly on the process of writing a summary. The percentages for

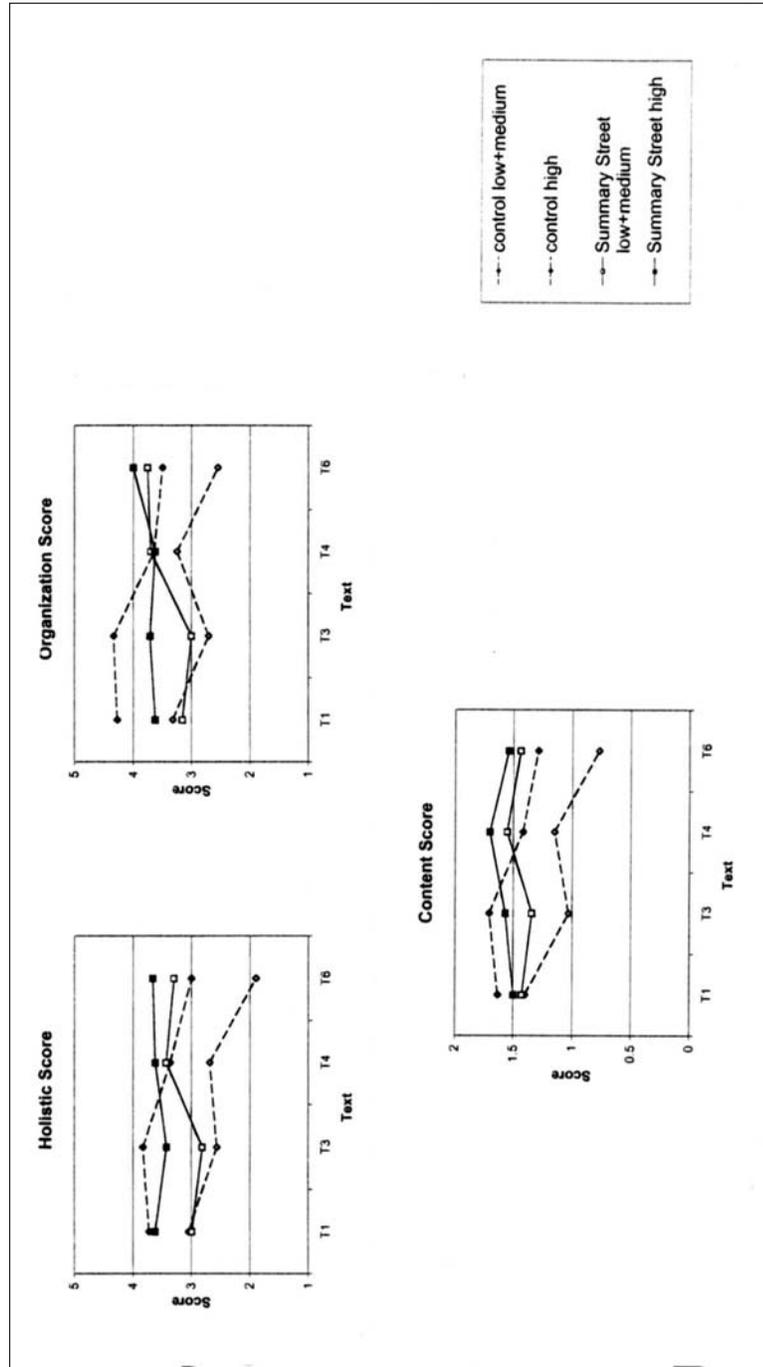


Figure 4. Measures of summary quality—holistic, content, and organization scores—as a function of condition, text, and ability level.

Summary Street[®] and control groups, respectively, are as follows: “cover the main points” (80% vs. 50%), “keep it short” (50% vs. 15%), “cover each paragraph” (25% vs. 15%), and “know the facts” (30% vs. 15%). Although these data are qualitative in nature, they nevertheless suggest that students who had used Summary Street[®] reflected more on what they were doing as they revised, and thus were not only able to write better summaries, but had become more aware of what kinds of processing are needed to do so. Perhaps this kind of insight contributed to their more thoughtful reading and better performance on the CSAP summary questions of the post-test.

DISCUSSION

The research reported here replicates the findings of Wade-Stein and Kintsch (2004), showing improvement in summaries written while supported by the content feedback that Summary Street[®] provides. It also extends the findings of the earlier study in several ways.

First, in contrast to the earlier study, the present study compared students who practiced writing summaries with and without feedback over an extended period of time: twice weekly sessions for 4 weeks. Overall scores on the comprehension test did not change from pre- to post-test for either the control or the Summary Street[®] group, nor did we find group differences for test items unrelated to the training (inference, vocabulary, fact finding, and other). However, we did observe a 25% gain among Summary Street[®] users for test items requiring a summary response. This finding supports our claim that summary practice with feedback that directs students to attend to relevant content can transfer to reading outside of the summary writing context. Simply practicing summary writing without such feedback is not sufficient. Our study lacks a non-summarizing control group—a design choice dictated by the availability of only four classes. However, there is no reason to expect that the performance of students with zero intervention would change in the course of 5 weeks. Students get relatively little practice in writing summaries in school, and when they do, teachers’ feedback usually addresses general problems directed to the class as a whole (e.g., cover the important information, say it in your own words, keep it short). The individualized, frequent feedback offered by Summary Street[®] thus fills a real need for the students. Moreover, this environment challenges students to deal with the content problems in their summaries on their own. Our observations in the classroom suggest that students work harder when using Summary Street[®], they tend to be more engaged in the task and take pride in their efforts, all of which seems to influence in a positive manner the way they subsequently read the expository texts they encountered on the CSAP post-test.

Second, the superior performance of students using Summary Street[®] on gist-level comprehension is also reflected in the better summaries they wrote. In blind scoring, summaries written with Summary Street[®] were judged to be superior on

several measures of writing quality: their overall quality, more complete content coverage, better organization, lack of low-level details, and stylistic quality.

There are several features of Summary Street[®] which interactively encourage students to write concise summaries that focus on important, relevant information in a text. The main content feedback, consisting of horizontal bars that correspond to each section heading, inform students when coverage is sufficient. The length indicator that is also part of the main feedback display makes students aware of how well their summaries fit within the length requirements. In the classroom we have often observed that students pass the content criteria early on and spend the rest of the time paring the length down to the allowable limit. By directing students to express important concepts in a limited space, students necessarily learn to eliminate overly detailed information and to express the content in their own words. In addition, the Relevance and Redundancy checks help users discriminate between more and less important information (Wade-Stein & Kintsch, 2004).

The significant difference in style scores between Summary Street[®] students and control students for the more difficult, later texts was unexpected (see Figure 3). There really is no mechanism in Summary Street[®] that would direct students to pay more attention to stylistic aspects of writing. However, as students are continuously revising their summaries to meet the content and length criteria, they might also catch and correct stylistic problems, thus leading to a more carefully written summary overall.

As expected, there were no differences between groups on the mechanics measure (Figure 3): although Summary Street[®] flags spelling errors on request, it does not target poor sentence structure or punctuation problems. There was relatively little evidence of plagiarism across the board. Summary Street[®] apparently does not encourage excessive sentence lifting, and students in both conditions generally tried to express the meaning in their own words (Figure 3).

A third point that deserves special emphasis is the fact that Summary Street[®] users were able to improve their summaries over the course of the 4-week intervention even as the texts they were working on became more difficult—in word length and with respect to structural as well as conceptual complexity. Earlier texts had one or two sections and were rated easy conceptually; later ones were four sections long and were considered moderately difficult. Thus, the patterns displayed in Figure 2 suggest that the feedback on request from Summary Street[®] enabled students to work successfully at a higher level with more demanding texts than would be possible without such support. This interpretation is reinforced by the data shown in Figure 4.

Finally, it is students of moderate-to-low ability levels who benefit the most from Summary Street[®]. The supportive context apparently makes it possible for these students to write summaries that are equivalent in terms of overall quality, content coverage, and organization to those of their higher achieving peers.

Although the present research has filled in some of the gaps in understanding what aspects of summary writing are benefitted by using Summary Street[®], there

are many issues we need to explore further. In particular, we plan to look at transfer effects in more detail, using post-summary questions and other tasks to assess students' content learning. Deep learning should be evaluated as well as gist recall: does Summary Street® enhance knowledge integration and consequently help students perform better on tasks that require inferential reasoning and problem solving, in comparison to reading and studying the material or even summarizing it without feedback? Longitudinal studies are currently underway to document possible growth in active, attentive comprehension and in summarizing and to see if the positive effects of long-term use of the tutor over the course of a semester or year persist when no feedback is provided. Another research direction concerns the development of metacognitive strategies: do students using Summary Street® gain a better sense of what to attend to during reading and do they improve their awareness of comprehension failures?

What the current study shows is that unsupported practice does not necessarily make perfect. Many studies have demonstrated that summary writing is a beneficial activity for students, for the effort to articulate one's understanding of a challenging text pays off in terms of better memory for the material. The extensive meta-analysis of writing-to-learn studies by Bangert-Downs, Hurley and Wilkinson (2004) largely supports this claim, though with the usual caveat: the effectiveness of writing to learn depends on many contextual factors, including, perhaps, feedback that leads students to reflect on their thinking, as Bransford, Brown, and Cocking (2000) maintain. Although the evidence on the role of feedback was not conclusive in the review cited above, the present study suggests that learning a complex skill like summarizing requires that the process be supported. Students' summaries and a crucial aspect of their comprehension improve when they are guided by a tutor that provides rapid, individualized feedback on demand about specific, yet typical problems in summary writing: getting the right content without wasting words. However, as Wade-Stein and Kintsch (2004) have pointed out, a good tutor also allows students to struggle and to find their own solutions without simply giving the right answer. Summary Street® is neither as intelligent nor as sensitive as a good human tutor, but it has the advantage that it can give students extensive, guided practice in open-ended writing way beyond what limited human resources can provide. As this study demonstrates, the computer tutor provides much needed help for a task that is difficult to support in the classroom: delivering individualized feedback on summaries of complex texts written in students' own words. The tutor is especially effective for medium- to low-achieving students, who rarely can get all the help they need. At the same time, it does not neglect the needs of higher achieving students.

ACKNOWLEDGMENTS

Members of the Summary Street® Research Group who also participated in this research are Nicole Davis, Walter Kintsch, Thomas Landauer, Mary Ann Stratton,

and Tim Weston. The authors are grateful for their participation and also gratefully acknowledge the contributions to this study of Lisa Kihn, Annie Kay Taylor, and of Ms. Stratton's 8th-grade language arts students at Century Middle School, Thornton, Colorado.

REFERENCES

- Angoff, W. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.; pp. 598-600). Washington, DC: American Council of Education.
- Bangert-Downs, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing to learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. National Research Council Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering and understanding. In J. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Cognitive development* (4th ed.; Vol. 3; pp. 515-629). New York: Wiley.
- Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*, 1-14.
- Cornoldi, C., & Oakhill, J. (Eds.). (1996). *Reading comprehension difficulties: Processes and intervention*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*, 1-53.
- Donovan, M. S., & Pellegrino, J. W. (Eds.). (2004). *Learning and instruction: A SERP research agenda*. Washington, DC: The National Academies Press.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kintsch, E., Steinhart, D., Stahl, G., LSA Research Group, Matthews, C., & Lamb, R. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments, 8*, 87-109.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris and S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71-92). Mahwah, NJ: Lawrence Erlbaum Associates.
- Landauer, T. K. (2002). On the computational basis of cognition: Arguments from LSA. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41; pp. 43-84). New York: Academic Press.
- Landauer, T. K., & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.
- Landauer, T. K., Laham, D., & Foltz, P. (2003). Automatic essay assessment. *Assessment in Education, 10*, 295-308.

- Landauer, T. K., McNamara, D. M., Dennis, S., & Kintsch, W. (in press). *Latent Semantic Analysis: A road to meaning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Landauer, T. K., & Psozka, J. (2000). Simulating text understanding for educational applications with latent semantic analysis: Introduction to LSA. *Interactive Learning Environments, 8*, 73-86.
- Miller, T. (2003). Essay assessment with Latent Semantic Analysis. *Journal of Educational Computing Research, 29*, 495-512.
- National Center for Educational Statistics (NCES). (2004). National Assessment of Educational Progress. *The nation's report card: Reading highlights 2003*. NCES: 2004-452. Washington, DC: US Department of Education.
- Paris, S. G., & Stahl, S. A. (Eds.) (2005). *Children's reading comprehension and assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.
- Shanteau, J. (2001). What does it mean when experts disagree? In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 229-244). Mahwah, NJ: Lawrence Erlbaum Associates.
- Trabasso, T., & Suh, S. Y. (1993). Using talk-aloud protocols to reveal inferences during comprehension of text. *Discourse Processes, 16*, 3-34.
- van den Brock, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah, NJ: Erlbaum.
- 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*, 333-362.

Direct reprint requests to:

Dr. Marita Franzke
Institute of Cognitive Science
University of Colorado
Boulder, CO 80309-0344
e-mail: mfranzke@psych.colorado.edu