A Contextualized Intervention for Community College Developmental Reading and Writing Students

Dolores Perin
Rachel Hare Bork
Community College Research Center
Teachers College, Columbia University

Stephen T. Peverly
Teachers College, Columbia University

Linda H. Mason
Pennsylvania State University, College of Education

Megan Vaselewski
Winston Preparatory School

January 2011

CCRC Working Paper No. 38

Address correspondence to:

Dolores Perin
Professor of Psychology and Education and
Senior Research Associate, Community College Research Center
Teachers College, Columbia University
525 West 120th Street, Box 174
New York, NY 10027
Email: perin@tc.edu

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through a grant (R305G060106) to the Community College Research Center, Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. Kristin Wisemiller and Amanda Rosenburg assisted with the development of the intervention and analysis of data. We are grateful to Susan De La Paz for her helpful comments on an earlier version of this paper.

Abstract

Instructors in community college developmental education programs are constantly seeking new ways to improve outcomes for their students, but, to date, there has been a shortage of empirical studies on the effectiveness of such efforts. The current study provides evidence on the potential efficacy of an approach to helping students develop an important academic skill, written summarization. In two experiments, a contextualized intervention was administered to developmental reading and writing students in two community colleges. The intervention was a 10-week curricular supplement that emphasized written summarization, as well as vocabulary knowledge, question generation, reading comprehension, and persuasive writing. The intervention was based on reading passages from science textbooks, with generic text from developmental education textbooks added in the second experiment.

In Experiment 1 (n = 322), greater gain was found for intervention than for comparison participants along three dimensions of written summarization: the proportion of main ideas from the source text included in the summary, accuracy, and word count (ES = 0.26–0.42). Experiment 2 (n = 246) set out to replicate and extend Experiment 1. Results were replicated for three of five summarization measures (ES = 0.36–0.70), but it was also found that intervention participants showed higher amounts of copying from the source text at posttest than the comparison group. In extending the intervention to a different text condition, it was found that students receiving science text outperformed students receiving generic text on the inclusion of main ideas, as well as on accuracy (ES = 0.32–0.33), providing moderate support for contextualization. Although summarization gains did not transfer to a standardized reading comprehension test in either experiment, the findings of this study suggest that the intervention had utility for academically underprepared postsecondary students.

Table of Contents

1. Introduction **1**
1.1 Written Summarization 2
1.2 Contextualization 4
1.3 Research Questions 5

2. General Method **6**
2.1 Overview 6
2.2 Intervention 7
2.3 Comparison Condition 12
2.4 Instructional Setting 13
2.5 Measures 14
2.6 Procedure 20
2.7 Analytic Strategy 21

3. Experiment 1 .... **22**
3.1 Participants 22
3.2 Results 23
3.3 Discussion 26

4. Experiment 2 .... **29**
4.1 Participants 29
4.2 Results 30
4.3 Comparison of Text Conditions 34
4.4 Discussion 34

5. General Discussion **36**
5.1 Level of Academic Preparedness 37
5.2 Potential Efficacy of Contextualization 38
5.3 Role of Science Knowledge and Science Interest 38
5.4 Paraphrasing Source Text when Summarizing 39
5.5 Transfer of Skill 40
5.6 Limitations 41

6. Conclusion **42**

References **43**

Appendix A: Scoring Criteria for Science Summarization Test, Accuracy, Conventions, and Paraphrasing Measures **50**

Appendix B: Science Interest Items **51**

1. Introduction

Large numbers of students who wish to earn a postsecondary credential enter the higher education system lacking the basic academic skills needed to learn at the postsecondary level. The consequences of this problem are both personal and social, as postsecondary education is associated with gainful employment (Kirst & Venezia, 2004). Among the approximately 45% of undergraduates in the United States who attend community college (American Association of Community Colleges, 2011), many are referred to developmental education (Bettinger & Long, 2005), which aims to prepare low-skilled students for postsecondary reading, writing, or mathematics requirements (Boylan, Bonham, & Tafari, 2005). For example, in a recent national sample of 51 community colleges, 31% to 39% of students were referred to developmental reading courses (Bailey, Jeong, & Cho, 2010).

Poor outcomes have been reported for developmental education (Bailey et al., 2010), but there is a lack of data on the effectiveness of specific instructional approaches for this population (H. M. Levin & Calcagno, 2008). Finding effective ways of preparing low-skilled students for postsecondary coursework has important implications for the future of community colleges. For instance, in a discussion of developmental education, Cohen and Brawer (2008) stated: "The overriding issue is whether community colleges can maintain their credibility as institutions of higher education even while they enroll increasingly less well-prepared students" (p. 281). The purpose of the research reported in this paper was to determine the potential effectiveness of an intervention for students in developmental reading and writing courses.

A problem identified in prior research was that students who have taken developmental courses often have difficulty with reading and writing requirements in college-level content courses (Perin & Charron, 2006). One task that underprepared postsecondary students find particularly difficult is the summarization of information (Caverly, Nicholson, & Radcliffe, 2004; Johns, 1985; Perin, Keselman, & Monopoli, 2003; Selinger, 1995; Spring & Prager, 1992). Summarizing information is a necessary component of various types of academic assignments and is frequently needed in postsecondary education (Bridgeman & Carlson, 1984). The primary focus of the present intervention was to help students improve their ability to write summaries contextualized in discipline-area text of the type they would encounter later in college-credit courses. In this paper, we begin with an overview of the demands of written summarization and then discuss the nature of contextualization. We then present data from two experiments conducted on the intervention. We conclude that the intervention has the potential to be useful in preparing developmental education students for specific academic tasks that they will face in college-credit courses.

1.1 Written Summarization

A summary is a distilled representation of the gist of source text (Hidi & Anderson, 1986; Rinehart & Thomas, 1993) and requires the identification of key ideas (Armbruster, Anderson, & Ostertag, 1987; A. L. Brown & Day, 1983). Summarization has been identified as a core academic skill: one of 10 core state literacy standards for college and career readiness in the United States is the ability to “determine central ideas or themes of a text and analyze their development; summarize the key supporting details and ideas” (National Governors’ Association & Council of Chief State School Officers [NGA/CSSO], 2010, p. 60).

Reading and writing, although usually taught as separate skills and in different developmental education courses, are often intertwined and mutually reinforcing in practice (Fitzgerald & Shanahan, 2000; Graham & Hebert, 2010). The task of written summarization occurs at the intersection of reading comprehension and writing (Mateos, Martin, Villalon, & Luna, 2008) and involves several broad processes: the creation of a mental representation of the key ideas in a text; the application of world knowledge in order to comprehend the text (Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005); and the ability both to extract only the information that is most important to the meaning (A. L. Brown & Day, 1983; Garner, 1985) and to produce a coherent piece of writing confined only to that information (Perin et al., 2003).

Johns (1985) reported that only 6% of a sample of low-skilled college students included at least two thirds of the main ideas in a history passage when writing a summary. Also, 88.9% of the summaries written by this group contained reproductions, which is defined as the copying of sentences from the source text. The summaries also contained much inaccurate information, such that 55.5% of the main ideas expressed in the summaries were scored as “distortions” (Johns, 1985, p. 506). Both learner and text variables can affect the quality of a written summary; Perin et al. (2003) found that community college developmental reading students who had less prior knowledge wrote shorter and less accurate summaries when the source text was relatively dense.

Effective summarization interventions have been reported for adolescents (Graham & Perin, 2007; Hare & Borchardt, 1984; Jitendra, Hoppes, & Xin, 2000; Reynolds & Perin, 2009; Rogevich & Perin, 2008), but few such studies have been conducted with postsecondary students. In a rare study with low-achieving college students, Selinger (1995) assessed the effects of five 1.25-hour intervention sessions involving the summarization of 1,200–1,400 word passages on business and psychology topics. The students’ summaries were collected and later returned with instructor comments. The teacher then lectured on the information that should be included in a summary and provided clues from the text on the thesis of the source text. Underlining and outlining were discussed, and students compared their own summaries with models. A randomly assigned control group read the same passages but only completed traditional reading comprehension exercises related to the material instead of receiving instruction. On a posttest summarization measure scored for the thesis, main ideas, and details from the source text, the treatment group showed statistically significantly larger gains than the control group.

The Selinger (1995) intervention evidently served to point students toward main ideas, suggesting that if students with literacy difficulties can be explicitly led to the location of main ideas in text, they may become more able to summarize the text. Low-skilled students may have problems reflecting on and monitoring their understanding of information as they read (Thiede, Griffin, Wiley, & Anderson, 2010), which may reflect a lack of awareness of the nature of main ideas themselves (Jitendra et al., 2000). It may not be so much that these students cannot identify a main idea as that they do not have a clear concept of what a main idea is. For this reason, pointing out the important ideas in text, a support provided in the present intervention, may address this weakness as low-skilled students attempt to generate a summary.

1.2 Contextualization

The present intervention was contextualized in reading passages drawn from science textbooks. The term “contextualization” has various meanings in the literature (Perin, 2011) but as used in this study refers to instruction embedded in content and applications that are relevant to students’ interests and goals (Johnson, 2002). What distinguishes contextualized instruction from traditional approaches is the sustained, systematic use of a single theme relevant to students’ academic and/or life goals. Any existing reading or writing practice can be contextualized, and in fact, the contextualization of instruction in specific discipline areas is the basis of content-area literacy taught in secondary education (McKenna & Robinson, 2009). In this approach, students learn reading and writing skills directly related to the genres and styles typical of particular disciplines.

Contextualization is thought by practitioners to address problems of limited transfer of skill (Carnine & Carnine, 2004; Tai & Rochford, 2007) and low motivation (Burgess, 2009; Dean & Dagostino, 2007). Generalization of skill may be facilitated through creating similarities between the contexts of instruction and application (Stone, Alfeld, Pearson, Lewis, & Jensen, 2006). For example, contextualized reading instruction focusing on science would address the characteristics of writing in science textbooks, including excessive use of difficult-to-comprehend, abstract language such as nominalized terms (Shanahan & Shanahan, 2008), and a lack of cohesion (Ozuru, Briner, Best, & McNamara, 2010), while instruction in the reading of a history text would include comparing multiple narratives for credibility and bias (Nokes, Dole, & Hacker, 2007). Improvement of reading and writing skills in developmental education courses based on sustained exposure to discipline-specific text may create conditions by which students can transfer the acquired skills to their college-credit courses in the disciplines in question.

Further, developmental education, which is often unpopular with students (Burgess, 2009), may be more motivating if it uses a disciplinary text that students know is typical of the texts that are assigned in concurrent or future discipline-area courses that they will have to pass to earn a postsecondary credential. However, basic skills instruction for low-achieving college students traditionally teaches literacy in a decontextualized and fragmented manner, which is rarely motivating (Grubb et al., 1999; Rose, 2005). In contrast, the innovation of contextualization is popular with educators who have used it (Baker, Hope, & Karandjeff, 2009), and there is a small amount of evidence suggesting that it is a promising approach for academically underprepared postsecondary students (Caverly et al., 2004; Martino, Norris, & Hoffman, 2001; Perin, 2011; Snyder, 2002).

1.3 Research Questions

This study investigated the relation between summarization skills and participation in a contextualized literacy intervention in a sample of community college developmental reading and writing students. The intervention consisted of a curricular supplement intended to strengthen students' ability to read the type of dense, informational text that they would later be expected to read in science and other disciplinary courses. The intervention provided students with practice in writing summaries and persuasive essays, defining vocabulary, formulating questions, and answering reading comprehension quizzes. Summarization was more strongly emphasized than the other skills in the intervention.

Two quasi-experimental studies, each lasting one college semester, were conducted with different samples. In the first experiment, the intervention was contextualized in science text, and outcomes were compared with those in a comparison group receiving the same developmental education curriculum but no intervention. The second experiment replicated and expanded upon the first experiment by randomizing participants to the science text condition or to a generic text condition. Performance on these groups was again compared to that of a comparison group. Both experiments controlled for science knowledge and interest, and student demographics.

The following questions were asked in Experiment 1: (1) Is participation in the intervention associated with better written summarization? (2) Does reading comprehension ability, as measured by the written summarization task, transfer to a standardized reading test? Experiment 2 asked: (1) Are results for the first experiment replicated with a different sample? (2) Does the impact of the intervention differ for the contextualized and generic conditions?

2. General Method

2.1 Overview

The commonalities between the two experiments are summarized in this section. The participants attended two community colleges, referred to as College 1 and College 2, which served a mixture of urban and suburban students in cities on the East and West Coasts, respectively. A purposive sample of 16 developmental education classrooms was recruited for each experiment; in both experiments, 12 of the classrooms received the intervention and four served as a comparison. The classes in each condition were divided evenly between the two sites. The instructors of these classes were recruited based on willingness to participate. All instructors had at least five years of experience in teaching developmental education. College administrators assigned the recruited instructors to experimental and comparison conditions based on instructors' stated interests.

In Experiment 1, the literacy practice was contextualized in science text; that is, all reading passages were from science textbooks and all literacy practice related to science topics. The science domain was selected because failure rates tend to be high in community college science courses, which impedes college graduation, and discussion with community college science instructors suggested that difficulty in reading course textbooks was an important factor in failure rates. In Experiment 2, a second condition, using generic text, was added, and participants were randomized to science and generic conditions within each experimental classroom. In the second experiment, the intervention was the same as in the first experiment and was identical for both text conditions, except that the vocabulary items and persuasive writing prompts were topic-specific.

In both experiments, all participants, both experimental and comparison, received the developmental education course curriculum as was typically delivered in the participating colleges, but whereas homework assignments were modified for the experimental participants such that completion of 10 intervention units was required over the semester, comparison participants were required to submit homework typical for the course. Participants in the experimental intervention completed the work outside of class. Ten percent of the course grade was awarded for submission of the completed units (10 units, one per week, 1% of course credit per unit). Homework in the comparison classrooms was given the amount of credit each instructor usually awarded. No financial remuneration was provided to participants, but instructors in both experimental and comparison classrooms were paid for their participation. The role of the instructors was to administer pre- and posttests, and in the experimental classrooms, to distribute and collect the weekly intervention units. The conduct of the study was facilitated by site coordinators who were developmental program directors at each site.

2.2 Intervention

The intervention, called the Content Comprehension Strategy Intervention (CCSI), was a pen-and-paper 10-unit curricular supplement that involved practice in written summarization, question formulation, vocabulary definition, persuasive writing, and the answering of multiple-choice reading comprehension questions. Each unit comprised a series of steps as follows: (1) activate prior knowledge by answering a question based on the title of the reading passage prior to reading it; (2) read a textbook passage; (3) check off items on a reading comprehension strategy checklist to indicate strategies used while reading; (4) select two words from a list of five technical terms (e.g., reactivity, anaerobic) or more general vocabulary (e.g., substance, matrix) from the text, look them up in a paper or online dictionary, copy the definition that fits with the passage, and then "write one sentence to explain the word to a friend;" (5) answer a self-efficacy question (based on Kitsantas & Zimmerman, 2009) on a 3-point scale concerning the vocabulary selected in the previous step ("If you see the word in a college textbook in the future, how sure are you that you will understand it immediately? Circle one number."); (6) prepare to write a summary of the reading passage by answering a series of questions focusing directly on main ideas explicitly stated in the passage; (7) write the summary; (8) answer a self-monitoring question about whether all the information from the prior answers to the main idea questions had been included in the summary, whether other ideas were included, whether the student's own words had been used, and whether the student had reread and corrected the summary; (9) formulate a question that an instructor might ask in class about the passage and then answer the question (based on Rosenshine, Meister, & Chapman, 1996); (10) take a 3-question multiple-choice reading comprehension quiz based on the passage; (11) write one or two paragraphs expressing an opinion on a controversy related to topic of the reading passage (based on De La Paz, 2005; Osborne, 2010; Schultz, 2003)—a simplification of an argumentative essay (see Ferretti, Lewis, & Andrews-Weckerly, 2009) tailored to students' abilities determined in pilot testing—which required the statement of the opinion, one reason for the opinion, and three supporting details; and finally, (12) judge the quality of the persuasive writing sample on a 6-point quality rubric used by the college where pilot testing had occurred but not by the colleges participating in the current study. At the end of each unit, students were asked to state how long they had taken to complete it.

All 12 steps were presented in the same order, and the instructions were formatted the same way in each unit. Only the reading passages and the content of the vocabulary questions and persuasive writing prompts varied across units. The intervention was used outside of the classroom independently in a self-directed, self-paced manner. Information from students indicated that each unit took between one and two hours to complete.

Text characteristics. Since an important dimension of contextualization is authenticity, the reading passages used in the two experiments were drawn intact from existing textbooks. The science passages, which were used in both experiments, were on anatomy and physiology. Initially, introductory community college-level text was intended for use throughout the science condition. However, pilot testing and discussion with instructors that took place while the intervention was being developed indicated that such text was too difficult, as students had little prior knowledge of the content, a problem also noted in the literature (Lei, Rhinehart, Howard, & Cho, 2010). Consequently, the 10 units in the science condition for both experiments were developed as five yoked pairs. The first unit of each pair presented a middle school level reading passage and the second unit of the pair used a passage on the same topic from the introductory level community college textbook used in the pilot study. Thus, the odd-numbered passages in the sequence of intervention units were from middle school textbooks and the even-numbered passages were on the community college level. It was expected that background knowledge and vocabulary would be developed using the easier text, which would then be applied to understanding the more difficult college-level text. The five yoked topics in the science condition were: matter and energy, atoms, the heart, blood, and respiratory system functions. The easier text was provided for review at the beginning of each even-numbered (college-level) unit.

Only college-level text was used in the generic condition (Experiment 2 only). The reading passages in this condition were on an assortment of themes and were selected from textbooks typical of, but not the same as, those used in the participants' developmental education classrooms. The generic reading passages concerned controversies on the following topics: genetic testing, entrepreneurship, censorship, drug addiction, the social consequences of air conditioning, the social role of news media, cosmetic surgery, participation of African Americans in baseball, youth hazing, and the founding of Liberia. The text was selected for its approximate match of word count with the science text.

The readability, word count, number of main ideas, and lexile scores for the intervention units in both conditions are summarized in Table 1. The number of main ideas was determined for each passage using a procedure described below in the section on measures. Mean Flesch-Kincaid readabilities, measured using the Microsoft Word program, were 8.99 (*SD* = 2.0) for the science text and 12.2 (*SD* = 1.34) for the generic text. The mean readability of the middle school science passages was 7.4 (*SD* = 1.27) and of the college level science text was 10.6 (*SD* = 1.03). The mean lexile scores, using the calculator at http://www.lexile.com/analyzer/, were 1077 (*SD* = 174.17) for the science text and 1268 (*SD* = 77.72) for the generic text, suggesting approximately ninth grade and college freshman levels, respectively. For the middle school level science text, the mean lexile score was 936 (*SD* = 20.74) with a mean of 1218 (*SD* = 134.61) for the college level science text, suggesting that the material was written at approximately the sixth and 12th grade levels, respectively (the lexile grade equivalents suggest higher reading grade levels than indicated by the Flesch-Kincaid readabilities). The mean science passage length was 610.50 words, (*SD* = 137.69) and the mean generic passage length was 607.7

words (*SD* = 119.16). The mean number of main ideas in the science text was 12.6 (*SD* = 2.84) and in the generic text was 9.1 (*SD* = 2.38).

Although attempts were made to equate text characteristics when selecting reading passages, there were statistically significant differences between the science and generic text in Flesch-Kincaid readability, lexile scores, and number of main ideas. Compared to the generic text, the science text had lower Flesch-Kincaid and lexile scores ($t = 4.25$, $df = 18$, $p = .000$ and $t = 3.17$, $df = 18$, $p = .005$, respectively), but more main ideas ($t = 2.99$, $df = 18$, $p = .008$). Word count for the two types of text did not differ ($t = .049$, ns).

Table 1
Experiments 1 and 2: Characteristics of Science and Generic Text

	Readability		Word Count		Number of Main Ideas		Lexile	
Unit	Science	Generic	Science	Generic	Science	Generic	Science	Generic
1	7.4	11.7	692	637	14	8	920	1370
2	9.1	12.7	576	562	18	8	1000	1290
3	8.1	10.6	477	451	14	6	950	1150
4	11.3	10.1	697	644	16	9	1300	1220
5	5.9	12.8	410	468	10	8	910	1300
6	10.7	13.0	565	551	11	7	1240	1220
7	6.5	11.8	691	691	12	12	940	1170
8	10.1	12.0	765	758	9	12	1200	1310
9	9.1	12.7	796	802	11	8	960	1380
10	11.7	14.9	436	513	11	13	1350	1270

Fidelity of implementation. In preparation for the intervention, project staff conducted orientation sessions with participating faculty and provided scripts for them to follow. Effective models of treatment integrity describe the quantity, quality, steps, and process of intervention delivery (Sanetti & Kratochwill, 2009). These dimensions were addressed with four measures of treatment integrity situated in the context of the intervention's service delivery settings. Intervention intensity, which is a compliance measure of the quantity of students' use of the intervention steps, was assessed first. The intervention consisted of 10 units, each of which comprised six tasks, for a total of 60 tasks across the full intervention. Not all students submitted all 10 units, and not all tasks were completed within each unit submitted, which would be expected from reports of sporadic attendance of community college students and high attrition rates (Crews & Aragon, 2007; Fike & Fike, 2008). (However, in the current data, there was not a statistically significant effect on the outcome variables of the number of units or steps completed.) Intensity was measured as both the number of students who submitted all 10 units and the mean number of tasks completed. In the first of two experiments conducted in this study,of 322 students who completed the pre- and posttests, 34% (n = 110) submitted all 10 intervention units. Among students who submitted all 10 units, the mean number of tasks completed was 53.53 (SD = 7.92), or 89% of the maximum 60 intervention tasks.

The second measure of treatment integrity asked whether students understood key tasks as intended by the intervention designers. To this end, three questions placed in three intervention units (for a total of nine questions) asked the student to state how he or she would explain to a friend how to do the task, for example:

> Please tell us: What is the best way to do Step 3, above? Write your answer so that a friend who is not doing these units would do all the parts of Step 3 in the same way you did it. We are interested in seeing *your own instructions*.

The questions concerned the vocabulary, written summarization, and question-formulation steps. Responses were scored on a 2-point scale as “correctly reported” or “incorrectly reported.” The three questions were placed in the seventh and ninth units only for a total of six questions. The responses were scored by a research assistant who was trained and supervised by the first author. Across the six questions, a range of 74% to 81% of responses were scored as correctly reported.

The third measure was a manipulation check (Sigall & Mills, 1998) that focused on students’ conceptualization of intervention tasks. Telephone interviews were conducted by an outside consultant toward the end of each intervention cycle with a purposive sample of 60 students in total. Instructors of each class receiving the intervention were asked by the first author to select four students to be interviewed: two students who were submitting completed intervention units on a regular basis and two who were not. However, students who were not submitting their work regularly were also frequently absent from class and for this reason were not available for interview. Therefore, all interviewees were students who were engaging consistently in the intervention.

Each respondent had a copy of the fifth intervention unit during the interview. The interviewer asked students five questions about several critical steps, for example: "Would you look at where it says Unit 5, Step 3? It's on page 5 of the unit. What do you have to do for that? How is that done?" The interview responses were scored by the interviewer as indicating correct versus incorrect understanding of the task. Results ranged from 66% to 89% correct understanding, with scores above 70% on each question with the exception of a question about the intervention step concerning vocabulary in the first cycle at College 2, which scored at 66%.

Interviewees were also asked (a) where they did the work for the unit, which was of interest because the intervention was a curricular supplement completed outside of the classroom; (b) how often they did the work (e.g., all at once or at different times during the week); (c) whether they received help with the work; and (d) what they saw as the purpose of the intervention. These four manipulation checks ranged from 56% to 100%. The only scores below 70% were for the item referring to how often the work was done at both sites (56% and 60%) and for the item regarding the purpose of the intervention at College 1 (61%). All interview responses were scored by the interviewer after extensive discussion with the first author. Although the first author worked closely with the scorers of the second and third treatment integrity measure, reliability was not measured.

Fourth, project staff communicated weekly with site coordinators at the two colleges to determine whether the instructors were administering the intervention as required; that is, site coordinators were asked if instructors were following the project scripts and were managing distribution and collection of the intervention units but were not helping students with the work. The site coordinators had daily contact with the instructors and met with them regularly to monitor project conduct. Based on information received from the site coordinators, all instructors administered the intervention as instructed.

2.3 Comparison Condition

Students in the comparison group followed the same classroom curriculum as the participants of the intervention but took only the pre- and posttests and received homework related to course material instead of the CCSI. The instructors and site coordinators verified that none of the homework or classroom instruction related to the subject matter in the intervention and that the comparison group did not engage in any special activities over and above regular course requirements.

2.4 Instructional Setting

Classroom curriculum. At College 1, the participants were enrolled in a developmental reading course and at College 2 they were enrolled in a developmental writing course. The developmental education program at each institution used a single syllabus and textbook and taught to departmental learning objectives. The textbooks used in both courses presented reading passages taken from larger texts in the humanities, literature, social science, and the sciences.

The course objectives at each college included aims that were related to the intervention under consideration in the current study in important ways. Objectives at College 1 included the use of context clues and word analysis to derive meaning from text and improve personal vocabulary; text comprehension through previewing, determining main ideas and supporting details; identification of author's purpose, mood, tone, point of view, assumptions, and intended audience; identification of figurative language and other literary devices; and understanding elements of fiction. College 2's learning objectives included use of elements of grammar such as sentence structure, verb and pronoun use, and subject-verb agreement; punctuation; spelling; word usage; writing paragraphs containing topic sentences and supporting detail; and reading at a ninth-grade level (the course integrated reading comprehension in the teaching of writing). Despite the differing curricular emphases, reading comprehension played a central role in the courses at both sites. Also, interviews with instructors indicated that written summarization, a major focus of the intervention, was occasionally assigned in class at both colleges.

College demographics. The demographics of the full student population at the participating colleges were as follows (demographics of the study population itself are provided when discussing experiment 1): At College 1, on the East Coast, mean age was 29 years, median 23; 61% were female; 22% Hispanic, 16% Black, 5% Asian, 48% White; 38% were enrolled full time; data on primary language were not collected. At

College 2, on the West Coast, 35% were aged under 20 years and 32% were aged 20–24 years; 55% were female, 31% were Latino, 6% Black, 16% Asian, and 34% White; and 71% were full time students. At College 2, 77% spoke English and 9% Spanish as a primary language. (College 1 was not able to provide information on primary language.)

2.5 Measures

Five dependent variables on written summarization were derived from a researcher-designed Science Summarization Test, and the Nelson-Denny Reading Test (J. I. Brown, Fishco, & Hanna, 1993) was administered as a transfer task. To control for preexisting knowledge of, and interest in, reading about science, two pretest measures—the Science Knowledge Test and the Science Interest Inventory—were also developed by the researchers. Several student background variables provided by the colleges were also used.

Science Summarization Test. The Science Summarization Test was a 30-minute task in which participants read a passage drawn from an introductory college anatomy and physiology textbook and wrote a summary with the text present. The instructions, which were provided in writing and read aloud by the instructor, asked the students to write a one- or two-paragraph summary that contained the important information in the passage (Armbruster et al., 1987). The instructions defined a summary as “a statement mostly in your own words that contains the important information in the passage.” Alternate forms A and B were developed, and pre- and posttest administration was counterbalanced to avoid text-specific effects. Form A consisted of 447 words, Flesch-Kincaid readability was 11.5, and lexile score was 1300. Form B contained 453 words, Flesch-Kincaid readability was 13.8, and lexile score was 1370. The topic of Form A was the nervous system and the topic of Form B was homeostasis. None of the information presented in these two reading passages overlapped with material presented in the text used in the intervention.

The five dependent variables obtained from the test were the proportion (i.e., percentage) of main ideas from the source text (Perin et al., 2003), the accuracy of information of the ideas expressed in the summary (Frey, Fisher, & Hernandez, 2003), word count, conventions (grammar, punctuation, and spelling), and the ability to paraphrase rather than copy information from the source (Keck, 2006). Although derived from the same task, most of the correlations among the scores in the two samples were weak, suggesting that discrete phenomena were being measured. There were only two correlations above $r = .4$ in the two samples: the proportion of main ideas and word count ($r = .48$, $p < .01$), and proportion of main ideas and accuracy ($r = .46$, $p < .01$), both of which occurred in the second experiment (see Table 2, which displays the correlations for both experiments).

In identifying the main ideas for the purpose of scoring, three criteria were applied: (1) the essential ideas in the source text that low skilled readers would be capable of understanding; (2) the learning objective, or the essential knowledge that should be learned given the intensity of information in the text, and (3) creating a narrative such that each main idea was linked to the one before it so that the summary tells a story, which reveals the organization of ideas in the text (Meyer, Middlemiss, & Theodorou, 2002).

The main ideas were identified by the third author and then checked by the first author and two research assistants who were state-certified high school teachers. It was determined that Form A contained 15 and Form B contained 17 main ideas. The main ideas in each student's summary were counted and then expressed as a proportion of the total number of main ideas in the source text. Proportions rather than raw scores were used because of the different number of main ideas in the source text for Forms A and B.

The accuracy of information in the written summary was measured on a 4-point scale, following a rubric reported by Frey et al. (2003). Word count was a simple count of the number of words written. Students' use of conventions, defined as grammar usage, punctuation, and spelling, was measured on a 4-point scale, also using Frey et al.'s (2003) rubric. Finally, the paraphrasing measure was a 2-point scale on the extent to which the summary was written in the student's own words (Perin et al., 2003), or "restating the ideas of a given excerpt without borrowing too liberally from the language of the original" (Keck, 2006, p. 262). The scoring criteria for the accuracy, conventions, and paraphrasing measures are provided in Appendix A.

Table 2
Experiments 1 and 2: Intercorrelations Among Pretest Measures

Variable (Experiment 1)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Nelson Denny								
1. Total scale score	1	.32***	.11*	.17***	--	.16**	.40***	.07
Science Summarization								
2. Prop. of main ideas		1	.18**	.22***	−.19**	.16**	.17**	−.02
3. Number of words			1	.13**	−.18***	−.06	.12*	.11*
4. Accuracy				1	−.36***	−.29***	.15**	−.04
5. Paraphrasing					1	.07	--	.02
6. Conventions						1	.09	−.15**
Background knowledge								
7. Science knowledge							1	.18***
8. Science Interest Inventory								1
Variable (Experiment 2)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Nelson Denny								
1. Total Scale Score	1	.19**	.15*	.11	.17*	.17*	.41**	.01
Science Summarization								
2. Prop. of Main Ideas		1	.48**	.46**	−.14*	−.03	.16*	.02
3. Number of Words			1	.05	−.03	−.13*	.12	.05
4. Accuracy				1	.24**	.12	.04	−.02
5. Paraphrasing					1	.06	.10	.02
6. Conventions						1	.08	−.09
Background Knowledge								
7. Science Knowledge							1	.32**
8. Science Interest Inventory								1

Note. * p < .05. ** p < .01. *** *p* < .001.

A research assistant experienced in writing assessment but unfamiliar with the goals of the project scored a random sample of 25% of the written summaries. Interrater reliabilities were $r = 0.92$ for proportion of main ideas and $r = 0.96$ for word count. Inter-scorer agreements were 90% for accuracy, 85% for conventions, and 83% for paraphrasing.

Nelson-Denny Reading Test. The Nelson-Denny Reading Test (J. I. Brown et al., 1993) consists of two subtests: a 15-minute, 80-item multiple-choice vocabulary subtest and a 20-minute reading comprehension subtest with 38 multiple-choice factual and inferential questions based on seven reading passages on a wide variety of topics. Scores on the two subtests are summarized in a total score. Scaled scores were derived using tables in the test manual, taking the first year of college as the reference. Because the vocabulary and reading comprehension subtest scores were highly correlated with each other and with the total score, only the total score was used. Internal consistency reliabilities reported in the manual for the test's alternate Forms G and H are 0.89 for vocabulary, 0.81 for reading comprehension, and 0.90 for the total score. Form G was administered at pretest and Form H at posttest. Test validity is not documented, but the measure has face validity for measuring general reading skills and identifying reading difficulties (Murray & Smith, 1998).

Science Knowledge Test. Prior knowledge about the content of material being read facilitates text comprehension (Carr & Thompson, 1996; Meyer & Rice, 1984). Background knowledge of content-specific information was tested with a researcher-developed Science Knowledge Test based on prior research (O'Reilly & McNamara, 2007). Unlike O'Reilly and McNamara, who used some questions from state standards, all questions in the new assessment were taken directly from the content units. All questions focused on knowledge relevant to the same domain as the passages. Three four-response, multiple-choice questions were initially developed from each of the 10 reading units for a total of 30 questions.

The questions were reviewed for coherence and suitability by an English professor with 10 years of experience in community college teaching. Following revision, two female adults (ages 22 and 24) with community college associate degrees completed the 30-question assessment. Each adult scored 78%. Interestingly, answers reflected differences in knowledge of biological science (nursing major) and earth science (education major). Following a discussion with the two adults, one question was eliminated from each three-question unit set, leaving 20 questions. The wording of the questions was also revised based on feedback from the two adults.

A descriptive analysis was completed following the administration of the test with 765 study participants, which refers to all participants who took the test during the pretest administration in Experiment 1 or Experiment 2, irrespective of whether they completed the intervention or the posttest. Results indicated that assessment scores were normally distributed with $M = 10.76$, $SD = 2.75$. Total items completed correctly ranged from 3–18 correct responses; no student scored 0–2 or 19–20 correct items. Visual inspection of the four-response items indicated that, overall, the response items ("A," "B," "C", and "D") were deemed acceptable by at least one respondent.

Univariate analysis with Bonferroni adjustment for comparisons was conducted to determine potential differences in effects for groups. Significance was set at $p = .01$. Significance was not found for assigned group, native language, or major. Males, however, performed significantly better than females (mean difference 0.72). In post hoc comparisons for race, only "Whites" when compared to "Other" (mean difference 1.45) was significant. The internal consistency (Cronbach's α) of the Science Knowledge Test was 0.63.

Science Interest Inventory. An assessment, based on the theoretical framework of *The Motivation for Reading Questionnaire* (*MRQ*, Wigfield & Guthrie, 1997) was developed to evaluate students' interest in science. As noted by Wigfield and Guthrie, researchers have established that interest in a topic affects comprehension positively, even when prior knowledge and intelligence are controlled. Although the *MRQ* was developed to assess students' motivation for reading, the framework of the questionnaire has been adapted and used previously in intervention research for story and persuasive writing (e.g., Harris, Graham, & Mason, 2006) and for expository reading comprehension plus informative writing (e.g., Mason, 2004).

The 4-choice Likert-type scale format of the newly developed 10-item Science Interest Inventory (SII) is similar to that of the *MRQ.* Students are asked to read a statement and respond by placing an "X" mark on the response that best tells how they feel. Points on the scale range from 1 (very different from me) to 4 (a lot like me). A practice sample item is provided prior to the first item. The student's response (e.g., 1, 2, 3, or 4) represents the score for the item. Three negatively worded items are inverted when scored (e.g., a score of 4 becomes a score of 1). The score for the scale is the average of the 10 items.

In the *MRQ*, interest is examined in four aspects: reading curiosity, which reflects the desire to learn; reading involvement, which reflects the pleasure gained when learning something of interest; the importance of reading, or subjective task value; and work avoidance, which reflects what students do not like about a topic. Items for the SII were developed to capture these aspects of interest (see Appendix B). For the purpose of the current research, *MRQ* questions or items related to interest were modified to reflect: (a) the topic of science, (b) the age of participants, and (c) the context of a community college.

Results of the SII, administered only at pretest to be used as a covariate, were used to examine the underlying dimensions of the items. First, intercorrelation between the items was tested. Findings indicated significant correlations for all items with the exception of one item, item 9: *I find it difficult to understand information about physical science such as atomic structure.* This item, however, did have significant correlations with six of the nine items. Results also indicated no multicollinearity among items.

It was assumed that the interest items would load on a single construct. Using students' responses, a principal components analysis was used to determine if the scale was unidimensional. The analysis generated the factor matrix with squared multiple correlations and two factors with eigenvalues greater than 1.0. These two factors accounted for 54% of the variance. The scree plot of eigenvalues indicated a relatively stable plateau after the second factor; therefore, a two-factor solution was rotated using the oblique solution. Results of rotation indicated that eight items had a pattern matrix loading greater than 0.40 on the first factor and four items had a factor structure loading greater than 0.40 on the second factor. Two of the items double-loaded on two factors. The items for the first factor appear to assess students' interest in science. The items in the second factor appear to assess the perception of science as a difficult subject. The internal consistency of the Science Interest Inventory (KR-20) was 0.53.

Student background variables. The colleges provided data on participants' age, gender, race/ethnicity, full- versus part-time college enrollment, and number of prior developmental education credits (these data are provided when discussing experiment 1).

2.6 Procedure

In each experiment, data were collected over 11 weeks of one semester, including both pre- and posttesting and completion of the intervention, at the two community colleges. Early in the semester, each participating instructor introduced the project, recruited students, and obtained signed consent according to script. The pretest was administered two weeks into the semester in all participating classrooms. In the experimental classrooms, immediately after the pretest, the first intervention unit was distributed for students' independent use over the coming week. For each subsequent week, the instructor collected the previous week's unit and distributed the next one. In the 11th week, when the 10th unit had been collected, the posttest was administered in both experimental and comparison classrooms.

The tests were administered by the instructors following project-developed scripts. All instructors had experience in administering tests. Project staff met with the site coordinator and instructors in the experimental condition to orient them to the testing and intervention procedures. The testing and instructional materials, and related directions and scripts, were sent by project staff to the sites for administration and distribution.

Prior to distribution, the intervention units were labeled by project staff with participants' names to ensure proper distribution and tracking. This was especially important in the second experiment when two different text conditions were administered. To facilitate the distribution and return process, each set of intervention units was printed on a different color of paper, with the two conditions printed in the same color. Instructors simply had to hand a unit to each student using the attached label. The site coordinator collected the completed materials from the instructors each week and sent them back to the project.

2.7 Analytic Strategy

Since the intervention required independent practice outside of the classroom and did not involve classroom instruction, the unit of analysis was the student, not the classroom. To assess pre–post gain in the intervention versus comparison group, the post-scores on the five dependent variables from the Science Summarization Test and the Nelson-Denny Reading Test (the transfer measure) were compared between groups using OLS regression with an analysis of covariance (ANCOVA), controlling for pretest scores. Science knowledge, science interest, and student background variables were also included in the analyses if they were found to be significantly correlated with the outcome measure.

Five analyses were performed for the dependent summarization variables: (1) the proportion of main ideas from the source text that were included in the summary, (2) word count, (3) the accuracy of information in the summary, (4) writing conventions, and (5) the extent to which information from the source text was paraphrased rather than copied. All scores were *z*-scores ($M = 0$, $SD = 1$). Another analysis was conducted on the Nelson-Denny total scores, using scale scores transformed from raw scores using tables provided by the publisher ($M = 200$, $SD = 25$).

Step 1 of each model adjusted for all background variables (science knowledge, science interest, and student background variables) found in prescreening to be related to the dependent variable, site of data collection, and pretest score. Step 2 introduced group status (1 = intervention; 0 = comparison) to determine whether the posttest scores varied by group, controlling for the scores used in Step 1. The regression weights are measures of effect size in predicting standardized posttest scores from group, standardized pretest scores, site of data collection, and background characteristics. Standardized beta weights were used as measures of effect size. Since the paraphrasing score was on a 2-point scale, this variable was analyzed using logistic regression; however, the method was identical to the OLS regression framework used in the other analyses.

3. Experiment 1

3.1 Participants

The initial sample for the first experiment consisted of 463 students, 35% from College 1 (n = 164) and 65% from College 2 (n = 299). The higher number of participants at College 2 was due to larger class sizes. The final sample consisted of 322 students (70% of the initial sample) who took both the pre- and posttests. Students were purposively assigned to two conditions, intervention and comparison. The numbers varied slightly by measure because some participants did not wish to take both tests at the pre- or posttest point. Among participants receiving the intervention, 245 took Nelson Denny pre- and posttests and 232 took Science Summarization pre- and posttests. In the comparison group, 72 took the Nelson Denny pre- and posttests and 67 took the Science Summarization pre- and posttests. Attrition between the pre- and posttesting appeared to be a random factor; univariate analyses of the pretest scores of the participants who did versus did not take the posttest ("completers" vs. "non-completers") indicated no statistically significant group differences. On the proportion of main ideas in the written summary on the Science Summarization Test, the mean score for completers and non-completers was the same (42% of main ideas, SD = 0.20 and SD = 0.21 respectively). On the Nelson-Denny total score, the mean for completers was 181.9 (SD = 23.52) and the mean for non-completers was 185.7 (SD = 18.27).

Of the final sample of 322, 33% (n = 107) were from College 1 and 67% (n = 215) were from College 2. The mean age was 19.71 years (SD = 4.75), and 67% were aged 18 years and younger; this was essentially a sample of older adolescents who had recently completed secondary education. Fifty-five percent of the sample were female; 37% were Hispanic, 33% White, 10% Asian, 9% Black, and 11% were other ethnicities. Sixty-eight percent were full-time students. At College 2, the native language was English for 72% and Spanish for 13% of the participants. Data on native language were not available at College 1, although the site coordinator stated that most participants were native speakers of English. Also, at both institutions, English language proficiency, as indicated by completion of English as a Second Language (ESL) courses or otherwise assessed by a college advisor, was a prerequisite for enrollment in developmental education. The large majority (93%) of students had no prior enrollment in a developmental reading or writing course, mostly because they had just entered the college.

Student background characteristics are shown in Table 3. Compared to the demographics of the college, summarized earlier, the study sample was younger and had greater racial/ethnic minority representation, although the gender breakdown was similar. Also, the proportion of study participants attending full time was closer to the general population of College 2 than College 1. Further, at College 1, slightly fewer of the study sample spoke English as a primary language than the general college population.

Table 3
Experiments 1 and 2: Participant Background Characteristics: Percentage of Sample

	Experiment 1 (*n* = 322)			Experiment 2 (*n* = 246)			
Variable	**Total Sample (*n* = 322)**	**Intervention (*n* = 249)**	**Comparison (*n* = 73)**	**Total Sample (*n* = 246)**	**Science (*n* = 97)**	**Generic (*n* = 97)**	**Comparison (*n* = 52)**
Race/ethnicity							
White	33.1	34.5	28.2	21.1	26.8	17.5	17.3
Black	9.4	9.2	9.9	20.3	18.6	22.7	19.2
Hispanic	36.6	33.3	47.9	34.1	28.9	35.1	42.3
Asian	10.3	12.4	2.8	14.6	12.4	18.6	11.5
Other	10.6	10.4	11.3	9.8	13.4	6.2	9.6
Gender							
Female	54.7	55.0	53.5	55.3	55.7	56.7	51.9
Male	45.3	45.0	46.5	44.7	44.3	43.3	48.1
Status							
Part time	31.6	30.1	36.6	40.0	41.2	29.2	57.7
Full time	68.4	69.9	63.4	60.0	58.8	70.8	42.3
Age							
18 & younger	66.6	65.5	70.4				
19 & older	33.4	34.5	29.6				
19 & younger				57.3	58.8	61.9	46.2
20 & older				42.7	41.2	38.1	53.8
Prior remedial credits							
0 credits	93.1	92.2	95.8	67.1	62.9	68.0	73.1
1 or more credits	6.9	7.8	4.2	32.9	37.1	32.0	26.9

3.2 Results

Descriptive statistics. A descriptive analysis of the two continuous summarization variables, proportion of main ideas, and word count was conducted using the data of all participants who completed the Science Summarization pretest (*n* = 429). The scores were normally distributed as follows: proportion of main ideas, *M* = 0.42, *SD* = 0.21, range 0–0.93; word count: *M* = 108.66, *SD* = 34.36, range 23–227. Univariate analyses were conducted as a function of gender, race/ethnicity, age, prior remedial credits, and part-time versus full-time attendance. Bonferroni adjustments were made and statistical significance was set at 0.01. For the conventions measure, students with zero remedial credits scored better than students with one or more remedial credits, mean difference 0.67 ($p < .01$), and students 18 years old and younger scored better than students 19 and older, mean difference 0.49 ($p < .001$).

The unadjusted pre- and posttest means and standard deviations for the five written summarization outcome variables and the Nelson Denny total reading score are displayed in Table 4. Sample sizes were slightly lower for the summarization measures because some students who took the Nelson-Denny posttest chose not to take the summarization posttest.

On the Science Knowledge Test (maximum score = 20), the mean score was 10.51 (*SD* = 2.56). On the Science Interest Inventory (maximum score = 40), the mean score was 26.65 (*SD* = 5.4). Table 2 shows the correlations among the pretest variables. There were no statistically significant group differences on the pretest summarization or reading measures, or on the science knowledge, science interest, or student background variables.

Prescreening of the relationship between the background variables and the outcome measures indicated statistically significant relationships between age and the conventions score; gender and the proportion of main ideas and word count scores; race/ethnicity and the Nelson Denny total scaled score; prior remedial credits and the Nelson Denny total scaled score and conventions score; science knowledge and the Nelson Denny total scaled score and proportion of main ideas and paraphrasing scores; and Science Interest and the Nelson Denny total scaled score and conventions scores.

Regression analyses. The results of the OLS regressions with ANCOVA were as follows. For the written summarization measure, post-scores were compared for the intervention and comparison groups on main ideas, word count, accuracy of information, conventions, and amount of paraphrasing, controlling for pretest score, site, and

Table 4
Experiment 1: Unadjusted Pre- and Posttest Scores

	Total Sample				Intervention Group				Comparison Group			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
Measure	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Sci. Sum. Test												
Prop. of main id.	0.42	0.21	0.48	0.21	0.43	0.21	0.50	0.21	0.42	0.20	0.43	0.22
Word count	108.94	33.86	111.62	35.17	109.30	34.78	115.10	34.40	107.69	30.69	99.57	35.39
Accuracy	2.92	0.79	2.94	0.73	2.88	0.82	2.98	0.72	3.04	0.68	2.82	0.74
Paraphrasing	0.76	0.43	0.67	0.47	0.74	0.44	0.64	0.48	0.84	0.37	0.76	0.43
Conventions	2.92	0.80	2.90	0.78	2.90	0.80	2.88	0.80	3.00	0.82	2.97	0.72
ND Reading Test												
Total scale score	185.66	18.27	184.31	20.67	185.89	18.64	183.65	21.22	184.90	17.03	186.56	18.62

Note. Sci. Sum.Test = Science Summarization Test; Prop. of main id. = Proportion of main ideas. Proportion of main ideas scores are in proportion form, counterbalanced. Maximum values for science summarization variables: accuracy = 4, conventions = 4, paraphrasing = 1. ND Reading Test = Nelson Denny reading test. Sample sizes vary based on group and assessment measure. Total sample (n = 317, n = 299); intervention group (n = 245, n = 232); comparison group (n = 72, n = 67).

background variables (science knowledge, science interest, and student characteristics). Table 5 is a summary of hierarchical regression analyses predicting intervention participants' posttest scores on main ideas, word count, accuracy and conventions, controlling for pretest scores, site of data-collection, and science knowledge, science interest, and student background variables.

Students who participated in the intervention included one third of a standard deviation more main ideas than the comparison group (ES = 0.34, $p < .01$). The intervention students included two fifths of a standard deviation more words than the comparison group (ES = 0.42, $p < .01$). The intervention group's posttest accuracy scores were one quarter of a standard deviation higher than those of the comparison group, controlling for pretest scores and site of data collection (ES = 0.26, $p < .05$).

Receiving the intervention was not a statistically significant predictor of conventions post-scores. For the paraphrasing variable, the overall model fit of the predictors (pretest score, site, science knowledge, science interest, and intervention condition) was very weak (−2 Log Likelihood = 347.59). The model correctly classified 68.6% of the cases but did not significantly predict group membership.
On the transfer measure, receiving the intervention was not a statistically significant predictor of posttest Nelson-Denny total scaled scores. However, there was a relation between the science knowledge and Nelson Denny scores. A 1-point increase on the Science Knowledge Test was associated with an approximately one and one half higher standard deviation Nelson Denny posttest score (ES = 1.45, $p < .001$).

3.3 Discussion

The experiment found promising results regarding the literacy skills of a low-achieving community college population. Controlling for science knowledge, science interest, site of administration, and student background variables, participants who received the intervention gained more than the comparison group on inclusion of main ideas, accuracy of information, and word count in summaries of dense, expository text on science topics, with effect sizes of 0.26 to 0.42.

There is not a robust body of intervention research on low-skilled adults to use in order to evaluate the size of these effects; only two studies were found (Friend, 2001;

Table 5
Experiment 1: Summary of Hierarchical Regression Analyses Predicting Intervention Participants' Science Summarization Posttest Scores, Controlling for Pretest Scores, Site of Intervention, and Student Background Characteristics (*n* = 294)

	Proportion of main ideas			Number of Words			Accuracy			Conventions		
Variables	***B***	***SE B***	**β**	***B***	***SE B***	**β**	***B***	***SE B***	**β**	***B***	***SE B***	**β**
Step 1												
Pretest	0.30***	0.05	0.31	0.25***	0.06	0.25	0.22***	0.06	0.22	0.35***	0.06	0.35
Site	0.28**	0.11	0.14	0.08	0.12	0.04	−0.20	0.12	−0.09	0.01	0.12	0.01
SK	0.04	0.02	0.09									
Female	0.38***	0.11	0.19	0.38***	0.11	0.19						
18 and younger										0.11	0.12	0.05
0 remedial credits										0.61**	0.24	0.15
SII										-0.01	0.01	-0.04
Step 2												
Pretest	0.30***	0.05	0.31	0.25***	0.06	0.25	0.23***	0.06	0.23	0.35***	0.06	0.35
Site	0.29**	0.11	0.14	0.08	0.12	0.04	−0.20	0.12	−0.10	0.01	0.12	0.01
SK	0.03	0.02	0.08									
Female	0.37***	0.11	0.19	0.38***	0.11	0.19						
18 and younger										0.11	0.12	0.05
0 remedial credits										0.61**	0.24	0.15
SII										−0.01	0.01	−0.04
Intervention	**0.34****	**0.13**	**0.14**	**0.42****	**0.13**	**0.17**	**0.26***	**0.14**	**0.11**	**−0.03**	**0.13**	**−0.01**

Note. ΔR^2 = 0.18 for Step 1 for proportion of main ideas ($p < .001$), ΔR^2 = 0.10 for Step 1 for number of words ($p < .001$), ΔR^2 = 0.05 for Step 1 for accuracy ($p < .001$), and ΔR^2 = 0.18 for Step 1 for conventions ($p < .001$); ΔR^2 = 0.02 for Step 2 for proportion of main ideas ($p < .01$), ΔR^2 = 0.03 for Step 2 for number of words ($p < .01$), ΔR^2 = 0.01 for Step 2 for accuracy ($p < .05$), and ΔR^2 = 0.00 for Step 2 for conventions *ns*; ΔR^2 = 0.001 for Step 3 for conventions *ns*; R^2 = 0.21 for Step 2 for proportion of main ideas ($p < .001$), R^2 = 0.13 for Step 2 for number of words ($p < .001$), R^2 = 0.06 for Step 2 for accuracy score ($p < .001$), and R^2 = 0.18 for Step 3 for conventions ($p < .001$). Students not receiving intervention are the comparison group. SII = Science Interest Inventory. SK = Science Knowledge Test. Female students are compared to male students (comparison group). Site compares students at College 1 to College 2 (comparison group). Students with zero previous remedial credits are compared to students with one or more previous remedial credits (comparison group). Students 18 years old and younger are compared to students 19 years and older (comparison group).

** $p < .01$. *** $p < .001$.

Selinger, 1995). Effect sizes of *d* = 0.56 (Friend, 2001) and 0.31 (Selinger, 1995) were calculated. Research with adolescents has shown effect sizes for summarization of *d* = 0.57 and *d* = 0.77 (Reynolds & Perin, 2009), and a meta-analysis found a mean weighted effect size of 0.82 for summarization (Graham & Perin, 2007). Since so few effect sizes on written summarization with college populations have been reported, it is useful also to consider the few effect sizes available in reading intervention research with low-skilled adults; effect sizes of *d* = 0.30 to *d* = 0.92 have been reported in these studies (Caverly et al., 2004; Hart & Speece, 1998; Snyder, 2002; Spring & Prager, 1992). The effect sizes found in the present study are low to moderate compared to the prior effect sizes.

On the main idea measure, unadjusted means increased seven percentage points for the intervention group but only one percentage point for the comparison group. Word count increased approximately 6% for the intervention group but decreased by 5% in the comparison group. Similarly, while accuracy increased in the treatment group, it decreased in the comparison group. Thus effect sizes for word count and accuracy may be attributable to declines in the comparison group. Both groups showed both poor performance and little pre–post change on both measures of the conventions of written English and the extent to which the source material was copied directly from the source text when writing in the summary.

The findings corroborate earlier studies (Johns, 1985; Perin et al., 2003; Selinger, 1995) that show underprepared college students' considerable difficulty in writing summaries. Although statistically significant gains were obtained in the present research, it should be noted that after one semester in a developmental education course that was supplemented with an intervention emphasizing written summarization, on average, the students included in a written summary only half of the key ideas in a reading passage from an introductory college science textbook. However, the results are encouraging in that they suggest that improvement is possible among this at-risk population.

The outcome measure was contextualized in science, as was the intervention. In considering the positive findings for the proportion of main ideas, accuracy, and word count, a question arose as to whether the results could be attributed to the intervention in general or to the contextualization of the literacy practice in the specific domain. A second experiment was conducted in order to replicate findings and to include an additional text condition in order to investigate the possible effects of contextualization.

4. Experiment 2

The data for the second experiment were collected in the semester immediately following the first, with a different cohort of students enrolled in the same developmental education courses. The intervention, design, and setting were identical except that a different text condition was added to the treatment, creating a contrast between contextualization in science text and generic text (see the General Method section above for details). Thus, while there were two groups, science and comparison, in Experiment 1, there were three groups, science, generic, and comparison, in Experiment 2.

4.1 Participants

A total of 365 students formed the initial sample for the second experiment, with 35% from College 1 (n = 128) and 65% from College 2 (n = 237), similar to the 33% and 67% from the two colleges in the first experiment. The total number of participants in the second experiment was smaller than for the first experiment because the data were collected in the spring semester, when enrollments tend to be lower. As before, the higher number of participants at College 2 was due to larger class size. The final sample consisted of 246 students (67% of the initial sample, similar to the 70% in the first experiment) defined as those who took both pre- and posttests. As in the earlier experiment, pretest scores between the completers and non-completers showed no statistically significant differences. On the proportion of main ideas in the written summary on the Science Summarization Test, the mean score for completers and non-completers was the same (39% of main ideas, SD = 0.21 and SD = 0.18 respectively). On the Nelson-Denny total pretest score, the mean for completers was 183.68 (SD = 21.08) and the mean score for non-completers was 180.98 (SD = 21.11).

As in the first experiment, 16 classrooms participated, with 12 forming the experimental group and four the comparison group. Within the experimental classrooms, the students were randomized to science and generic text conditions. In the science text condition, 85 students took the Nelson Denny pre- and posttests and 82 students took the Science Summarization pre- and posttests. In the generic text condition, 85 students took the Nelson Denny pre- and posttests and 77 students took the Science Summarization pre- and posttests. In the comparison group, 49 students took the Nelson Denny pre- and posttests and 40 students took the Science Summarization pre- and posttests. The pretest scores of completers and non-completers showed no statistically significant differences.

The mean age of participants in the second experiment was slightly higher than that in the first at 21.43 years (*SD* = 6.02), with 57% aged 19 years and younger, compared to a mean age of 19.71 and 67% aged 18 and younger in the first experiment. As can be seen in Table 3, compared to the sample in the first experiment, the participants of the second experiment showed the same representation of females, similar Hispanic representation, and showed somewhat fewer full-time students, a smaller proportion of White students, and higher proportions of Asian and Black students.

4.2 Results

Descriptive statistics. Using the data of all participants ($n = 337$) who completed the pretest, a descriptive analysis of the proportion of main ideas, and word count measures was performed. The scores were normally distributed as follows: proportion of main ideas, $M = 0.39$, $SD = 0.19$, range 0–0.94; word count: $M = 104.46$, $SD = 34.15$, range 11–224 words. Univariate analyses were conducted as a function of gender, race/ethnicity, age, and part-time versus full-time attendance. Bonferroni adjustments were made, with statistical significance set at 0.01. For the paraphrasing measure, students with zero remedial credits scored better than students with one or more remedial credits ($p < .01$).

Table 6 shows the unadjusted pre- and post-scores for the five summarization variables and the Nelson Denny total scaled score. On the Science Knowledge Test (maximum score = 20) the mean score was 10.82 ($SD = 2.90$), and on the Science Interest Inventory (maximum score = 40) the mean score was 27.06 ($SD = 5.45$). These scores were similar to those found in the first experiment (the correlations among the pretest variables for both experiments are shown above in Table 2). As in the first experiment, the three groups (science, generic, comparison) did not differ statistically on the pretest

Table 6
Experiment 2: Unadjusted Pre- and Posttest Scores

	Total Sample				**Science Text Condition**				**Generic Text Condition**				**Comparison Group**			
	Pretest		**Posttest**		**Pretest**		**Posttest**		**Pretest**		**Posttest**		**Pretest**		**Posttest**	
Measure	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Sci. Sum. Test																
Prop. of main Id.	0.39	0.18	0.46	0.23	0.41	0.19	0.52	0.22	0.40	0.19	0.48	0.22	0.33	0.15	0.32	0.21
Word count	106.05	35.62	108.60	41.65	108.21	34.22	117.43	39.72	109.57	38.17	115.03	35.62	95.40	31.85	79.84	43.78
Accuracy	2.90	0.66	3.06	0.63	2.99	0.68	3.22	0.63	2.90	0.61	3.01	0.54	2.74	0.69	2.86	0.71
Paraphrasing	0.83	0.38	0.69	0.46	0.78	0.42	0.57	0.50	0.83	0.38	0.74	0.44	0.91	0.29	0.84	0.37
Conventions	2.86	0.83	2.82	0.84	2.77	0.88	2.77	0.79	2.85	0.81	2.66	0.84	3.07	0.71	3.22	0.82
ND Reading Test																
Total scale score	181.11	21.22	185.87	22.36	183.19	20.89	185.81	21.31	182.09	21.55	188.19	21.85	176.06	21.11	181.96	24.83

Note. Sci. Sum. Test = Science Summarization Test; Prop. of main id. = Proportion of main ideas. The proportion of main ideas scores are in proportion form, counterbalanced. Maximum values for Science Summarization Variables: Accuracy = 4, Conventions = 4, Paraphrasing = 1. ND Reading Test = Nelson Denny Reading Test Sample sizes vary based on group and assessment measure. Total Sample ND Reading Test (n = 219); Science Text Condition (n = 85); Generic Text Condition (n = 85); Comparison Group (n = 49). Total Sample Science Summarization Test (n = 199); Science Text Condition (n = 82); Generic Text Condition; (n = 77); Comparison Group (n = 40).

summarization or reading variables, or on the science knowledge, science interest, or student background measures.

When the background variables and outcome measures were prescreened, statistically significant relationships were found between prior remedial credits and the Nelson Denny total scaled score and conventions scores; race/ethnicity and the proportion of main ideas and conventions scores; full-time status and proportion of main ideas and word count scores; citizenship status and conventions scores; gender and proportion of main ideas scores; science knowledge and Nelson Denny total scaled scores; and Science Interest and conventions scores.

Regression analyses. Participation in the intervention was associated with gain on several written summarization variables, but not on the Nelson Denny transfer measure. Table 7 summarizes the hierarchical regression analyses predicting post-scores on the Science Summarization Test, controlling for pre-scores, site, science knowledge, science interest, and student demographics and academic background. Students receiving the science text included over one half of one standard deviation more main ideas from the source text in their summaries, compared to students in the comparison group (ES = 0.62, $p < .001$). Similarly, students receiving the generic text also included more main ideas compared to students in the comparison group (ES = 0.36, $p < .05$).

Participants in both intervention conditions showed greater gain in word count than the comparison group. Compared to the comparison group, science text participants wrote 0.70 *SD* more words ($p < .001$), and generic text students wrote 0.62 *SD* more words than the comparison group ($p < .001$). The science text group's posttest accuracy scores were 0.44 *SD* higher than those of the comparison group ($p < .05$). However, gain on posttest accuracy was not different in the generic text and comparison groups. The outcome for the conventions measure was similar for the experimental and comparison groups.

On the paraphrasing variable, the overall model fit of the predictors (pretest score, site, science knowledge, science interest, and intervention condition) was weak (−2 Log Likelihood = 227.679), but was statistically reliable in distinguishing posttest scores $\chi^2 = 8.46$, $p < .05$). The model correctly classified 73.4% of the cases. The comparison group was four times more likely to summarize the source text in their own words than the

Table 7
Experiment 2: Intervention vs. Comparison Group. Summary of Hierarchical Regression Analyses Predicting Intervention Participants' Science Summarization Posttest Scores, Controlling for Pretest Scores and Student Background Characteristics (*n* = 205)

	Proportion of Main Ideas			Number of Words			Accuracy			Conventions		
Variables	***B***	***SE B***	**β**	***B***	***SE B***	**β**	***B***	***SE B***	**β**	***B***	***SE B***	**β**
Step 1												
Pretest	0.45***	0.06	0.45	0.41***	0.06	0.41	0.32***	0.07	0.32	0.35***	0.06	0.35
Site	0.33*	0.14	0.16	0.46***	0.13	0.23	0.22	0.14	0.11	0.15	0.17	0.07
SII										−0.03**	0.01	−0.17
Female	0.15	0.12	0.07									
0 remedial credits										0.45**	0.15	0.22
Asian	0.50**	0.19	0.17							−0.52**	0.20	−0.18
White	0.13	0.17	0.05							−0.08	0.17	−0.03
Black	0.00	0.18	0.00							−0.40*	0.18	−0.17
Other race	−0.58**	0.22	-0.17							0.25	0.23	0.07
Full time	0.21	0.13	0.10	0.21	0.14	0.10						
US Citizen										0.29	0.16	0.12
Step 2												
Pretest	0.42***	0.06	0.42	0.37***	0.06	0.37	0.29***	0.07	0.29	0.33***	0.06	0.33
Site	0.33*	0.14	0.16	0.46***	0.13	0.23	0.20	0.14	0.10	0.16	0.17	0.08
SII										−0.03**	0.01	-0.16
Female	0.15	0.12	0.07									
0 remedial credits										0.44**	0.15	0.21
Asian	0.47**	0.18	0.16							−0.48*	0.20	-0.17
White	0.03	0.16	0.01							−0.08	0.17	-0.04
Black	−0.04	0.17	−0.02							−0.38*	0.18	-0.16
Other race	−0.63**	0.21	−0.19							0.21	0.23	0.06
Full time	0.15	0.13	0.08	0.11	0.13	0.06						
US citizen										0.28	0.16	0.12
Science	**0.62*** **	**0.16**	**0.31**	**0.70*** **	**0.16**	**0.34**	**0.44***	**0.18**	**0.22**	**−0.24**	**0.17**	**−0.12**
Generic	**0.36***	**0.16**	**0.18**	**0.62*** **	**0.16**	**0.30**	**0.16**	**0.18**	**0.08**	**−0.39***	**0.17**	**−0.19**

Note. ΔR^2 = 0.34 for Step 1 for proportion of main ideas ($p < .001$), ΔR^2 = 0.25 for Step 1 for number of words ($p < .001$), ΔR^2 = 0.12 for Step 1 for accuracy ($p < .001$), and ΔR^2 = 0.31 for Step 1 for conventions ($p < .001$); ΔR^2 = 0.05 for Step 2 for proportion of main ideas ($p < .001$), ΔR^2 = 0.07 for Step 2 for number of words ($p < .001$), ΔR^2 = 0.03 for Step 2 for accuracy ($p < .05$), and ΔR^2 = 0.02 for Step 2 for conventions *ns*; ΔR^2 = 0.1 for Step 2 for conventions ($p < .001$); R^2 = 0.39 for Step 2 for proportion of main ideas ($p < .001$), R^2 = 0.32 for Step 2 for number of words ($p < .001$), R^2 = 0.16 for Step 2 for accuracy score ($p < .001$), and R^2 = 0.33 for Step 3 for conventions ($p < .001$). Hispanic students are the uncoded comparison group for the White, Asian, Black, and Other variables. SII = Science Interest Inventory. Female students are compared to male students (comparison group). Site compares students at site 2 to site 3 (comparison group). Students with 0 previous remedial credits are compared to students with one or more previous remedial credits (comparison group). Full-time students are compared to part-time students (comparison group). US citizens are compared to non-US citizens (comparison group). Female students are compared to male students (comparison group).

* $p < .05$. ** $p < .01$. *** $p < .001$.

science group. Thus, the posttest summaries of the science group showed a greater increase in the amount of copying from the source text than the comparison group. There was no difference between the comparison and generic conditions on this variable.

4.3 Comparison of Text Conditions

Table 8 presents the predictability, for the science and generic text conditions, of posttest scores on the proportion of main ideas, number of words, accuracy, and conventions in the written summaries, controlling for pretest scores. The comparison group was removed from this analysis in order to compare the science and generic groups directly. As in Experiment 1, all pre- and posttest measures were standardized ($M = 0$, $SD = 1$). Step 1 includes the pretest score, site, and all background characteristics found to have a statistically significant relationship with the dependent variable. Differences between the science and generic conditions were found for main ideas and accuracy but not for word count, conventions or paraphrasing. Controlling for pretest, site, and background characteristics, students receiving the science text included one-third of a standard deviation more main ideas in their summaries than students in the generic text condition (ES = 0.32, $p < .05$). The science text group's posttest accuracy scores were 0.33 *SD* higher than those of the generic text group, controlling for pretest scores, site, and background characteristics ($p < .05$). As previously noted, the science group was more likely than either the generic or comparison group to copy from the source text rather than paraphrase the material in their own words.

4.4 Discussion

For the most part, the second experiment replicated the first. The intervention group, both science and generic conditions, showed greater gain than the comparison group on the proportion of main ideas and word count of the written summaries (ES = 0.36–0.70, compared to ES = 0.34–0.42 in Experiment 1). The science group also showed greater gain than the comparison group on the accuracy of the summary. Neither treatment condition differed from the comparison group on written English language conventions, again replicating the first experiment. However, unlike the first experiment, the science group showed greater gain on the paraphrasing measure than the comparison

Table 8
Experiment 2: Science vs. Generic Text Condition. Summary of Hierarchical Regression Analyses Predicting Science Intervention Participants' Science Summarization Posttest Scores, Controlling for Pretest Scores, Site, and Student Background Characteristics
(*n* = 151)

	Proportion of Main Ideas			Number of Words			Accuracy			Conventions		
Variables	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Step 1												
Pretest	0.42***	0.07	0.42	0.38***	0.08	0.38	0.21*	0.08	0.20	0.27***	0.08	0.27
Site	0.36*	0.17	0.18	0.44**	0.16	0.22	0.22	0.17	0.11	0.05	0.16	0.03
SII							−0.03	0.02	−0.14	−0.06***	0.01	−0.30
19 and younger										0.20	0.16	0.10
Asian	0.41	0.22	0.15									
White	−0.05	0.19	−0.02									
Black	−0.18	0.21	−0.07									
Other race	−0.93***	0.27	−0.27									
Full time	0.27	0.16	0.13	0.22	0.16	0.11						
US citizen										0.19	0.18	0.08
Step 2												
Pretest	0.41***	0.07	0.41	0.38***	0.08	0.38	0.20*	0.07	0.29	0.27***	0.07	0.27
Site	0.36*	0.16	0.18	0.44**	0.16	0.21	0.21	0.17	0.10	0.05	0.16	0.02
SII							−0.03	0.02	−0.15	−0.06***	0.01	−0.30
19 and younger										0.19	0.16	0.10
Asian	0.43*	0.22	0.16									
White	−0.13	0.19	−0.05									
Black	−0.19	0.20	−0.08									
Other race	−1.00***	0.27	−0.28									
Full time	0.31*	0.15	0.15	0.23	0.16	0.11						
US citizen										0.22	0.18	0.09
Science	**0.32***	**0.14**	**0.16**	**0.11**	**0.15**	**0.05**	**0.33***	**0.16**	**0.16**	**0.19**	**0.15**	**0.10**

Note. ΔR^2 = 0.32 for Step 1 for proportion of main ideas ($p < .001$), ΔR^2 = 0.21 for Step 1 for number of words ($p < .001$), ΔR^2 = 0.09 for Step 1 for accuracy ($p < .01$), and ΔR^2 = 0.22 for Step 1 for conventions ($p < .001$); ΔR^2 = 0.02 for Step 2 for proportion of main ideas ($p < .05$), ΔR^2 = 0.003 for Step 2 for number of words *ns*, ΔR^2 = 0.03 for Step 2 for accuracy ($p < .05$), and ΔR^2 = 0.01 for Step 2 for conventions *ns*; ΔR^2 = 0.1 for Step 2 for conventions ($p < .001$); R^2 = 0.35 for Step 2 for proportion of main ideas ($p < .001$), R^2 = 0.21 for Step 2 for number of words ($p < .001$), R^2 = 0.12 for Step 2 for accuracy score ($p < .001$), and R^2 = 0.23 for Step 3 for conventions ($p < .001$). Students receiving the generic CCSI intervention are the comparison group. Hispanic students are the uncoded comparison group for the White, Asian, Black, and Other variables. SII = Science Interest Inventory. Female students are compared to male students (comparison group). Site compares students at site 2 to site 3 (comparison group). Students 19 years old and younger are compared to students 20 years old and older (comparison group). Full-time students are compared to part-time students (comparison group). US citizens are compared to non-US citizens (comparison group).

* $p < .05$. ** $p < .01$. *** $p < .001$.

group, while the amount of gain on this measure between the generic and comparison groups did not differ.

Comparing the results of the first experiment, in which all treatment participants received science text, with the science condition in the second experiment, indicates that the second experiment replicated the first on four of five measures of written summarization (proportion of main ideas, word count, accuracy, and conventions) and produced different findings for the remaining variable, paraphrasing. Further, the two experiments showed the same results on the transfer measure, such that group membership was not a predictor of gain on the Nelson-Denny Reading Test.

Most of the effect sizes in both experiments were in the moderate range in comparison to other research on summarization and/or literacy interventions with adolescents and postsecondary students (see effect sizes for this work in the discussion of the first experiment, above). Two of the effect sizes were stronger, main ideas of the intervention versus comparison group (ES = 0.62) and word count of both science and generic versus comparison (ES = 0.70 and ES = 0.62, respectively). However, although these effects were stronger, they were still not as strong as the mean weighted effect size of $d = 0.82$ found in the meta-analysis of Graham and Perin (2007).

The second experiment found moderate support for the contextualization of the literacy intervention. Specifically, compared to the generic text condition, the science contextualization condition resulted in the inclusion of a greater proportion of main ideas and greater accuracy of information (ES = 0.32 and 0.33), but also four times more copying from the source text when writing a summary.

5. General Discussion

The present research begins to fill a gap in understanding the literacy skills of a large but overlooked at-risk population in the educational intervention literature, students who have graduated from secondary education but who enter postsecondary institutions with low reading and writing skills. Participation in a supplement to the developmental education curriculum that emphasized written summarization and also provided practice in vocabulary, question generation, reading comprehension questions and persuasive writing was associated with gain on several dimensions of written summarization. In two experiments, students receiving the intervention showed greater gain than a comparison group on the proportion of main ideas from source text, accuracy of information, and word count. In addition, contextualization of the intervention in science text was associated with superior performance on inclusion of main ideas and accuracy of written summaries on science topics. Despite some limitations discussed below, the research is important in suggesting that the literacy skills of community college developmental education students can improve.

5.1 Level of Academic Preparedness

The mean Nelson Denny Reading Test posttest scores of 184 and 186 for the samples in the two experiments were below the test mean of 200, although the participants had attended a developmental education course one level below the college level for one semester. Furthermore, even though participation in the intervention was associated with gain on a written summarization task, by the end of the intervention, participants were still missing many of the main ideas when they summarized text from an introductory level college science textbook. Notwithstanding any gain associated with the current intervention, this finding suggests that despite their placement into the highest level of developmental reading or writing, the students were far from ready for college level work.

As a measure of reading comprehension (Graham & Hebert, 2010), summarization is subject to the effects of prior knowledge (McKeown, Beck, Sinatra, & Loxterman, 1992). The science knowledge measure indicated that the students had limited background knowledge of the science topics they were being asked to summarize in the pre- and post-summarization tests. However, this does not seem to be the only explanation for the difficulty. Johns (1985) and Selinger (1995) also reported similar problems using generic text in earlier studies. Our results, in combination with the previous research suggest that while summarization enables learning from text (Armbruster et al., 1987), given that this skill is also vulnerable to the effects of prior knowledge, it is important to strengthen background knowledge at the same time as teaching summarization skills to underprepared students.

5.2 Potential Efficacy of Contextualization

Contextualization has some strong advocates (Baker et al., 2009; Johnson, 2002), but its benefits as an intervention have rarely been directly tested at any educational level, let alone in postsecondary settings (Perin, 2011). The current study provides some evidence for the potential efficacy of this approach in finding positive effects on several measures of written summarization. These results have direct implications for practices within developmental education, the purpose of which is to teach the basic academic skills needed to learn effectively from the college curriculum. Developmental reading and writing courses are typically designed as preparation for the first level of college English (Edgecombe, 2011; Perin & Charron, 2006), but developmental education students also need preparation in reading large amounts of dense, informational text, since this will be expected in disciplinary courses.

Anecdotal evidence from several studies conducted in community colleges by the first author indicates that developmental education instructors prefer to use generic material on the assumption that it will be more likely to transfer to a range of tasks, whereas contextualization is viewed as being too narrow an approach. However, the students receiving the generic condition in this study did not do as well on several measures of summarization as students who received text contextualized in the same subject area as the text used in the outcome measure. Nevertheless, the issue of contextualization needs further investigation, first in light of the theoretical debate regarding the extent to which instruction should be general versus narrow in order to promote the transfer of skill (Anderson, Reder, & Simon, 1996; Bransford, Brown, & Cocking, 2000), and second, because the current study used only one measure relating to contextualization. A fuller explanation would be gained by using both generic and contextualized outcome measures.

5.3 Role of Science Knowledge and Science Interest

Two researcher-developed measures were administered at pretest to control for science knowledge and science interest. Scores on these measures revealed that overall, the sample knew little of the content of the science intervention and had a low level of interest in the domain of science. On average, students answered 10 of 20 items correctly on the science knowledge measure and indicated an average score of 27 on a 40-point Likert-type scale of science interest. Science knowledge was significantly correlated with the Nelson-Denny total scaled score ($r = 0.4$, $p < .001$) and with several measures of written summarization, although at much lower levels ($r = 0.15–0.17$). The weak relation between prior knowledge and the researcher-developed summarization measure may be explained by a floor effect, as scores were low on both measures.

The science interest measure had a weak relationship with the written summarization measures and no statistically significant relationship with the Nelson-Denny Reading Test scores. However, despite students' generally low interest in science, they were able to benefit from an intervention using science text.

The study suffered from attrition, and although it is possible that low science interest might be partially responsible, no differences in any of the pretest scores, including science interest, were found between students who stopped participating and those who persevered. The problem of attrition appeared not to be related to experimental conduct and in fact is common in community colleges (Barnett, 2011; Glass & Oakley, 2003).

5.4 Paraphrasing Source Text when Summarizing

Substantial amounts of copying directly from the source text were observed in this study using the paraphrasing measure. The phenomenon of copying extensively when writing a summary has also been found in postsecondary English language learners of typical skill levels (Keck, 2006). It appeared that the current students improved in the ability to detect what was important in the source text, based on the improved scores on the proportion of main ideas from the text included in the summaries, but at the same time, in Experiment 2, they copied more. This result suggests that the students' writing skills did not keep up with their increase in sensitivity to important information in text. Thus, it appears that the students receiving the intervention started to see what was important but could not state it in their own words. Hidi and Anderson (1986) reported that children tended to copy word-for-word when summarizing text, but by approximately sixth grade, the summaries consisted of the writer's own words. The current data suggest that academically underprepared college students, and possibly English language learners (Keck, 2006), may not have made that transition. A direction for future research would be to investigate the relation between the paraphrasing measure and the accuracy of the summaries. Although the accuracy of the summaries increased from pre- to posttest, the level of accuracy remained somewhat low (e.g., a post-score mean of 2.99, *SD* 0.68 on a 4-point accuracy scale for participants in the science condition in the second experiment), even though the source text was present while the summaries were written.

Using a different scoring procedure to measure the accuracy of written summaries, Perin et al. (2003) excluded sentences largely or completely copied from the source text. When these sentences were discarded, two thirds of the sentences written were judged as accurate, suggesting higher levels of accuracy than in the current sample. The issue of plagiarism is routinely raised in preparing students for college level literacy demands (e.g., see McWhorter, 2010). Given the obvious importance of accuracy in academic writing, it would be useful to investigate the relation between this variable and the phenomenon of copying from text.

5.5 Transfer of Skill

The gains shown on the written summarization measure did not transfer to the measure of general reading skills, the Nelson Denny Reading Test. The groups did not differ in their gain on this measure, and in the group as a whole, pre–post gain was not statistically significant. This finding supports the possibility of the benefits of contextualization, since performance on a measure not related to the content of the intervention reading passages did not change. However, this finding may also reflect a tendency in the educational intervention literature such that greater gain is often shown on researcher-designed than standardized measures, possibly because the researcher measures are more closely associated with instructional variables (Rosenshine & Meister, 1994). There is little information in the literature about the relation between generic measures such as the Nelson Denny Reading Test and performance in college-level courses. However, note-taking research by Peverly and Sumowski (2011) with typically-performing four-year college students found Nelson-Denny scores to be related to performance on test items that required the learner to make inferences. Future research is needed to determine the predictiveness of both contextualized and generic measures that are both standardized and researcher-developed, and college achievement.

5.6 Limitations

The current study suggests that the intervention has potential efficacy, but conclusions remain tentative due to several methodological limitations. For example, although participants were randomized to intervention conditions in the second experiment, the intervention and comparison groups were purposive samples. College administrators assigned classes to experimental and comparison conditions based on local considerations beyond the scope of the project. Also, the students were randomized to intervention conditions within classrooms, and the extent of contamination (the possibility that experimental students may have shared intervention information with students in the comparison group) is unknown. Future research using this type of intervention design should collect teacher and student interview data to determine whether contamination may have occurred. Further, the comparison group was small; using a larger group may have allowed identification of stronger results if they did exist. Additionally, in the contextualization of the intervention, only one subject area, anatomy and physiology, was used, and so the possible benefits of contextualization in other content areas was not evaluated. In addition, the intervention did not include detailed feedback to students. Participants' low levels of knowledge and interest in the releveant subject area is another limitation. Stronger effects of contextualization may be found if knowledge and interest are higher, based on theories that this approach is beneficial because it makes learning relevant to students' interests and needs (Johnson, 2002).

Further, some of the positive findings relate to pre–post declines in the comparison group. This group showed a decline from pre- to posttest on word count in both experiments, and on accuracy in the first experiment. The declines may have been due to actual lowering of the skills tested in the measures or to lower motivation on the posttest. However, a decline in skill at posttest on the part of the comparison group would suggest that a supplement such as that tested in this study is important to help maintain students' skills as they proceed through a developmental education course, but the current data are not sufficient to support this possibility. In addition, the generic text in the second experiment had higher mean readability and lexile scores than the science text, which raises the possibility that the better outcome in the science versus generic text group may be related more to the ease of comprehension than to the intervention experience. The various limitations arising in this study suggest the need for further research on the effects of this intervention.

6. Conclusion

The sample of students participating in this study were mostly low-achieving older adolescents and young adults who were products of U.S. secondary education. Previous research suggests that community college developmental education instruction is limited in its effects. Students enter with enormous obstacles resulting from personal history and academic background (Cohen & Brawer, 2008; J. S. Levin, 2007), and the pedagogy used may not be compelling (Grubb et al., 1999). In this context, the positive results obtained in this research are notable. The findings indicate that, although the students were low-skilled, they could still improve in ways that carried potential to promote their future success in college-credit courses. The intervention included standard literacy skills that are taught but not necessarily learned across the grade levels. Continued practice in these skills, contextualized in disciplinary content, seems to be a promising direction toward the preparation of academically underprepared students for the reading and writing demands of the postsecondary curriculum.

References

American Association of Community Colleges. (2011). *2011 fact sheet: Building a nation of learners by advancing America's community colleges*. Washington, DC: Author. Retrieved from http://www.aacc.nche.edu/AboutCC/Documents/FactSheet2011.pdf

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

Armbruster, B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly, 22*(3), 331–346.

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*(2), 255–270.

Baker, E. D., Hope, L., & Karandjeff, K. (2009). *Contextualized teaching and learning: A faculty primer.* Sacramento, CA: Research and Planning Group for California Community Colleges, Center for Student Success. Retrieved from http://www.careerladdersproject.org/docs/CTL.pdf

Barnett, E. A. (2011). Validation experiences and persistence among community college students. *The Review of Higher Education, 34*(2), 193–230. doi: 10.1353/rhe.2010.0019

Bettinger, E. P., & Long, B. T. (2005). Remediation at the community college: Student participation and outcomes. *New Directions for Community Colleges, Spring*(129), 17–26.

Boylan, H. R., Bonham, B. S., & Tafari, G. N. (2005). Evaluating the outcomes of developmental education. *New Directions for Institutional Research, Spring*(125), 59–72.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press. Available from http://www.nap.edu/catalog.php?record_id=9853

Bridgeman, B., & Carlson, S. B. (1984). Survey of academic writing tasks. *Written Communication, 1*(2), 247–280. doi: 10.1177/0741088384001002004

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 1–14.

Brown, J. I., Fishco, V. V., & Hanna, G. S. (1993). *The Nelson-Denny Reading Test, Forms G and H.* Itasca, IL: Riverside/Houghton-Mifflin.

Burgess, M. L. (2009). Using WebCT as a supplemental tool to enhance critical thinking and engagement among developmental reading students. *Journal of College Reading and Learning, 39*(2), 9–33.

Carnine, L., & Carnine, D. (2004). The interaction of reading skills and science content knowledge when teaching struggling secondary students. *Reading & Writing Quarterly, 20*(2), 203–218.

Carr, S. C., & Thompson, B. (1996). The effects of prior knowledge and schema activation on the inferential reading comprehension of children with and without learning disabilities. *Learning Disability Quarterly, 19*(1), 48–61.

Caverly, D. C., Nicholson, S. A., & Radcliffe, R. (2004). The effectiveness of strategic reading instruction for college developmental readers. *Journal of College Reading and Learning, 35*(1), 25–49.

Cohen, A. M., & Brawer, F. B. (2008). *The American community college* (5th ed.). San Francisco, CA: Jossey-Bass.

Crews, D. M., & Aragon, S. R. (2007). Developmental education writing: Persistence and goal attainment among community college students. *Community College Journal of Research and Practice, 31*(8), 637–652. doi: 10.1080/10668920600859715

De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology, 97*(2), 139–156.

Dean, R. J., & Dagostino, L. (2007). Motivational factors affecting advanced literacy learning of community college students. *Community College Journal of Research and Practice, 31*(2), 149–161. doi: 10.1080/10668920600859657

Edgecombe, N. (2011). *Accelerating the academic achievement of students referred to developmental education* (CCRC Working Paper No. 30, Assessment of Evidence Series). New York, NY: Columbia University, Teachers College, Community College Research Center.

Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101*(3), 577–589.

Fike, D. S., & Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review, 36*(2), 68–88.

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist, 35*(1), 39–50.

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*(1), 53–80.

Frey, N., Fisher, D., & Hernandez, T. (2003). “What’s the gist?” Summary writing for struggling adolescent writers. *Voices from the Middle, 11*(2), 43–49.

Friend, R. (2001). Effects of strategy instruction on summary writing of college students. *Contemporary Educational Psychology, 26*, 3–24.

Garner, R. (1985). Text summarization deficiencies among older students: Awareness or production ability? *American Educational Research Journal, 22*(4), 549–560. doi: 10.3102/00028312022004549

Glass, J. C., Jr. , & Oakley, B. (2003). Attrition and retention among accounting majors in community colleges: Problems and possible remediation. *Community College Journal of Research and Practice, 27*(8), 679–698. doi: 10.1080/10668920390211075

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading* (A Carnegie Corporation Time to Act Report). Washington, DC: Alliance for Excellent Education. Retrieved from http://carnegie.org/fileadmin/Media/Publications/WritingToRead_01.pdf

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476.

Grubb, W. N. (with Worthen, H., Byrd, B., Webb, E., Badway, N., Case, C., . . . Villeneuve, J. C.). (1999). *Honored but invisible: An inside look at teaching in community colleges*. New York, NY: Routledge.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445–459.

Hare, V. C., & Borchardt, K. M. (1984). Direct instruction of summarization skills. *Reading Research Quarterly, 20*(1), 62–78.

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43*(2), 295–340.

Hart, E. R., & Speece, D. L. (1998). Reciprocal teaching goes to college: Effects for postsecondary students at risk for academic failure. *Journal of Educational Psychology, 90*(4), 670–681.

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research, 56*(4), 473–493.

Jitendra, A. K., Hoppes, M. K., & Xin, Y. P. (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.

Johns, A. M. (1985). Summary protocols of “underprepared” and “adept” university students: Replications and distortions of the original. *Language Learning, 35*(4), 495–517.

Johnson, E. B. (2002). *Contextual teaching and learning: What it is and why it’s here to stay*. Thousand Oaks, CA: Corwin Press.

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. doi: 10.1016/j.jslw.2006.09.006

Kirst, M. W., & Venezia, A. (Eds.). (2004). *From high school to college: Improving opportunities for success in postsecondary education*. San Francisco, CA: Jossey-Bass.

Kitsantas, A., & Zimmerman, B. J. (2009). College students’ homework and academic achievement: The mediating role of self-regulatory beliefs. *Metacognition and Learning, 4*(2), 97–110

Lei, S. A., Rhinehart, P. J., Howard, H. A., & Cho, J. K. (2010). Strategies for improving reading comprehension among college students. *Reading Improvement, 47*(1), 30–42.

Levin, H. M., & Calcagno, J. C. (2008). Remediation in the community college: An evaluator’s perspective. *Community College Review, 35*(3), 181–207.

Levin, J. S. (2007). *Nontraditional students and community colleges: The conflict of justice and neoliberalism*. New York, NY: Palgrave Macmillan.

Martino, N. L., Norris, J. A., & Hoffman, P. R. (2001). Reading comprehension instruction: Effects of two types. *Journal of Developmental Education, 25*(1), 2–10.

Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology, 96*(2), 283–296.

Mateos, M., Martìn, E., Villalòn, R., & Luna, M. (2008). Reading and writing to learn in secondary education: Online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing: An Interdisciplinary Journal, 21*(7), 675–697. doi: 10.1007/s11145-007-9086-6

McKenna, M. C., & Robinson, R. D. (2009). *Teaching through text: Reading and writing in the content areas* (5th ed.). Boston, MA: Pearson Education.

McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly, 27*(1), 78–93. doi: 10.2307/747834

McWhorter, K. T. (2010). *College reading and study skills* (11th ed.). Boston, MA: Longman.

Meyer, B. J. F., Middlemiss, W., Theodorou, E., Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Effects of structure strategy instruction delivered to fifth-grade children using the Internet with and without the aid of older adult tutors. *Journal of Educational Psychology, 94*(3), 486–519.

Meyer, B. J. F., & Rice, G. E. (1984). The structure of text. In D. Pearson (Ed.), *Handbook of reading research* (Vol. 1, pp. 319–351). White Plains, NY: Longman.

Murray, W. M., & Smith, D. K. (1998). Nelson-Denny Reading Test, Forms G and H. *Mental measurements yearbook* 13.

National Governors' Association, & Council of Chief State School Officers [NGA/CSSO]. (2010). *Common core state standards: English language arts and literacy in history/social studies, science, and technical subjects*. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology, 99*(3), 492–504.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes: A Multidisciplinary Journal, 43*(2), 121–152.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*(5977), 463–466. doi: 10.1126/science.1183944

Ozuru, Y., Briner, S., Best, R., & McNamara, D. S. (2010). Contributions of self-explanation to comprehension of high- and low-cohesion texts. *Discourse Processes: A Multidisciplinary Journal, 47*(8), 641–667.

Perin, D. (2011). Facilitating student learning through contextualization: A review of evidence. *Community College Review, 39*(3), 268–295. doi: 10.1177/0091552111416227

Perin, D., & Charron, K. (2006). “Lights just click on every day:” Academic preparedness and remediation in community colleges. In T. Bailey & V. S. Morest (Eds.), *Defending the community college equity agenda* (pp. 155–194). Baltimore, MD: Johns Hopkins University Press.

Perin, D., Keselman, A., & Monopoli, M. (2003). The academic writing of community college remedial students: Text and learner variables. *Higher Education, 45*(1), 19–42.

Peverly, S. T., & Sumowski, J. F. (2011).What variables predict quality of text notes and are text notes related to performance on different types of tests? *Applied Cognitive Psychology*. doi: 10.1002/acp.1802

Reynolds, G. A., & Perin, D. (2009). A comparison of text structure and self-regulated writing strategies for composing from sources by middle-school students. *Reading Psychology, 30*(3), 265–300. doi: 10.1080/02702710802411547

Rinehart, S. D., & Thomas, K. F. (1993). Summarization ability and text recall by novice studiers. *Reading Research and Instruction, 32*(Summer), 24–32.

Rogevich, M. E., & 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.

Rose, M. (2005). *Lives on the boundary: A moving account of the struggles and achievements of America’s educationally underprepared*. New York, NY: Penguin Books.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479–530. doi: 10.3102/00346543064004479

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*(2), 181–221.

Schultz, M. (2003). Analyzing argumentative strategies. *Teaching English in the Two-Year College, 31*(1), 77–80.

Selinger, B. M. (1995). Summarizing text: Developmental students demonstrate a successful method. *Journal of Developmental Education, 19*(2), 14–16, 18, 20.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.

Sigall, H., & Mills, J. (1998). Measures of independent variables and mediators are useful in social psychology experiments: But are they necessary? *Personality and Social Psychology Review, 2*(3), 218–226.

Snyder, V. (2002). The effect of course-based reading strategy training on the reading comprehension skills of developmental college students. *Research and Teaching in Developmental Education, 18*(2), 37–41.

Spring, C., & Prager, J. (1992). Teaching community-college students to follow the train of thought in expository texts. *Reading and Writing: An Interdisciplinary Journal, 4*(1), 33–54.

Stone, J. R., III, Alfeld, C., Pearson, D., Lewis, M. V., & Jensen, S. (2006). *Building academic skills in context: Testing the value of enhanced math learning in CTE* (Final study). Columbus, OH:: The Ohio State University, National Dissemination Center for Career and Technical Education. Retrieved from http://136.165.122.102/UserFiles/File/Math-in-CTE/MathLearningFinalStudy.pdf

Tai, E., & Rochford, R. A. (2007). Getting down to basics in Western Civilization: It's about time. *Community College Journal of Research and Practice, 31*(2), 103–116.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes: A Multidisciplinary Journal, 47*(4), 331–362.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology, 89*(3), 420–432.

Appendix A: Scoring Criteria for Science Summarization Test, Accuracy, Conventions, and Paraphrasing Measures

Accuracy (4-point scale)

1 = Most statements cite outside information or opinions and/or most of the statements are inaccurate and not verified by the text.

2 = Some statements cite outside information or opinions and/or some of the statements are inaccurate and not verified by the text.

3 = Most statements are accurate and verified by the text.

4 = All statements are accurate and verified by the text.

Conventions (4-point scale)

1 = 10+ errors in grammar, punctuation, or spelling; reader cannot follow what is written

2 = 7–9 errors in grammar, punctuation, or spelling

3 = 4–6 errors in grammar, punctuation, or spelling

4 = 0–3 errors in grammar, punctuation, or spelling

Paraphrasing (2-point scale)

0 = Contains at least three strings of words, each one consisting of at least five words, copied directly from the text.

1 = Mostly in student's own words; large majority not copied from the text.

Appendix B: Science Interest Items

Curiosity (3 items)

3. I like to read about new and unfamiliar science information. (Factor 1)

4. I enjoy learning about different science topics. (Factor 1)

8. I would like to know more about science topics such as biology. (Factor 1)

Involvement (3 items)

1. I like to learn about science topics that are hard and challenging. (Factors 1 & 2)

2. I do not like it when information about science makes me think. (Factor 2)

6. I like learning about science topics outside of work and college. (Factor 1)

Importance (2 items)

5. If the science topic is interesting, I do not care how much time it takes to learn about it. (Factor 1)

10. I personally believe that it is important to know about science topics such as biology. (Factor 1)

Avoidance (2 items)

7. I do not like science. (Factors 1 and 2)

9. I find it difficult to understand information about physical science such as atomic structure. (Factor 2)