Journal of Educational Psychology
2011, Vol. 103, No. 3, 578–593

0022-0663/11/$12.00 DOI: 10.1037/a0023930

The Effectiveness of a Question-Exploration Routine for Enhancing the Content Learning of Secondary Students

Janis A. Bulgren, Janet G. Marquis, B. Keith Lenz, and Donald D. Deshler
University of Kansas

Jean B. Schumaker
Edge Enterprises

The purpose of this study was to examine the effects of a question-exploration routine and an associated graphic organizer on students' ability to think about and answer complex questions. Participants were 116 students of diverse abilities in seven 7th grade classes. The effects of the routine were compared with the effects of a traditional lecture-discussion format using a counterbalanced design. The measure, composed of matching, multiple-choice, and short-answer questions, assessed students' knowledge and comprehension of facts, main ideas, and relationships that require higher order thinking. Overall, significant differences representing large to very large effect sizes were found between the total test scores of students in the 2 groups. Specifically, students taught using the question-exploration routine earned higher total test scores than did students taught using the lecture-discussion method. Similar differences were found for matching, multiple-choice, and short-answer items. The scores earned by subgroups of students (high achievers, average achievers, low achievers, and students with disabilities) followed the same general pattern. Students in all subgroups had more difficulty conveying main ideas in writing than in multiple-choice formats.

Keywords: questioning, graphic organizer, secondary content-area instruction, students with diverse abilities, students with disabilities

Supplemental materials: http://dx.doi.org/10.1037/a0023930.supp

Recent data show that students in American schools are not achieving as well as students in other nations. According to the Alliance for Excellent Education, students in other countries outperform students in the United States on cross-national assessments (Rothman, 2009). In fact, students in the United States ranked 24th of 29 countries in problem solving, and half of American students fell below the threshold of problem-solving skills necessary to meet emerging workforce demands (Organization for Economic Co-Operation and Development, 2007). As a result, many students are unable to succeed in core courses, pass state assessments, or achieve at levels required to compete in today's world (Conley, 2010; Heller & Greenleaf, 2007). These findings require a reassessment of educational goals and instructional procedures if students are to compete in the worldwide economy.

To compete in the worldwide economy, students need intellectual tools and learning strategies that allow them to acquire knowledge and think productively (Bransford, Brown, & Cocking, 2000). Instead of focusing on the relatively simple skills of reading, writing, and calculating, students are now being challenged to become lifelong learners who can think critically about and answer meaningful questions. According to Graesser, Baggett, and Williams (1996), such questions, if used correctly, can be fundamental guides for human reasoning. Answering meaningful questions requires that students not only know facts and concepts but that they can also make inferences and use information in new situations (W. Kintsch, 1994). The importance of this challenge is reiterated in national standards across the content areas, such as science and social studies, that stress the importance of answering meaningful questions and solving problems (International Reading Association, 2006; National Center for History in the Schools, 1996; National Council of Teachers of English, 1996; National Council of Teachers of Mathematics, 2000; National Research Council, 1996, 2000).

Unfortunately, some teachers may not have access to instructional procedures that can help students answer complex questions. As a result, these teachers are not adequately prepared to teach the higher order thinking skills required of students today (Raudenbush, Rowan, & Cheong, 1993). For example, research has shown that the questions commonly posed in classrooms are often shal-

This article was published Online First July 18, 2011.

Janis A. Bulgren, Center for Research on Learning, University of Kansas; Janet G. Marquis, Schiefelbusch Institute for Life Span, University of Kansas; B. Keith Lenz, Department of Special Education, University of Kansas; Donald D. Deshler, Center for Research on Learning and Special Education, University of Kansas; Jean B. Schumaker, Edge Enterprises, Lawrence, KS.

B. Keith Lenz is now a Senior Researcher at SRI International, Menlo Park, California.

This research was supported in part by Office of Special Education Programs Grant H023C50143-96 to the University of Kansas Center for Research on Learning. The author wishes to disclose that after this original research was conducted, a manual for teachers was written and is currently available for purchase.

Correspondence concerning this article should be addressed to Janis A. Bulgren, Center for Research on Learning, University of Kansas, J. R. Pearson Hall, 1122 West Campus Road, Lawrence, KS 66045. E-mail: jbulgren@ku.edu

low, rather than deep (Graesser et al., 1996), and some teachers focus on delivery of large amounts of content information versus instruction of higher order thinking (Bulgren et al., 2006; Torff, 2003).

An additional challenge that complicates the problem of teaching students to answer complex questions is the diversity among groups of students in many classrooms. Groups of students in typical classes often include students who represent various levels of achievement, including high achievers (HA), average achievers (AA), and low achievers (LA). In addition, classes contain students with disabilities (SWD), including students who have learning disabilities (LD) and students who have other health impairments (OHI). Recent reviews of trends indicate that students with disabilities, including those with LD, are expected to meet curriculum standards adopted by states and professional organizations. Furthermore, these students will likely be included in the same classrooms as students without disabilities for the majority of the school day (Swanson & Deshler, 2003). Therefore, it is critical to understand the learning characteristics and diverse needs of these students.

The need for instruction that considers students of diverse achievement and ability levels is highlighted by recent legislation and studies. To illustrate, recent national legislation in the Blueprint for Reform contains expectations that "every student graduates from high school well prepared for college and a career" (U.S. Department of Education, Office of Planning, Evaluation and Policy Development, 2010). Unfortunately, whereas other nations embarked on educational reforms late in the 20th century (Darling-Hammond, 2010), the United States has only recently responded, in part because it is falling behind other developed nations (Organization for Economic Co-Operation and Development, 2007; Rothman, 2009). As a result, national organizations, including Carnegie Corporation of New York, Alliance for Excellent Education, National Governors Association, and Institute of Education Sciences, have called for more attention to be directed toward interventions designed to help struggling adolescent learners.

A growing body of research demonstrates that learners respond differently to interventions designed to promote comprehension. These learners represent varying skills (McNamara, O'Reilly, Best, & Ozuri, 2006), knowledge (McNamara, 2001, 2004), dispositions (Azevedo & Cromley, 2004; Azevedo, Cromley, & Seibert, 2004; Greene & Azevedo, 2007), and achievement and ability levels (Bulgren, Deshler, Schumaker, Lenz, 2000; Bulgren, Lenz, Schumaker, Deshler, & Marquis, 2002). Furthermore, students often respond differently to different types of questions designed to assess comprehension (Campbell & Mayer, 2009; Davey & McBride, 1986; King, 1994). These findings are compatible with research indicating that struggling students, such as those with LD, often have difficulty answering questions that ask for more information than facts, details, or vocabulary definitions (Pressley et al., 1992; Vaughn, Gersten, & Chard, 2000) and may lack skills for processing and organizing information, making inferences, understanding relationships, and distinguishing main ideas from details (DiCecco & Gleason, 2002). In addition, students are increasingly being asked to respond on assessments in writing, and many students have difficulty writing responses (Bulgren, Marquis, Lenz, Schumaker, & Deshler, 2009).

To respond to this situation, teachers and students need support in several areas. First, teachers need access to instructional procedures that help students achieve at high levels of comprehension and thinking as they answer questions. Procedures must be research based and shown to help students with various abilities. This is particularly important because teachers generally do not adopt innovative instructional procedures unless those procedures have been shown to benefit a wide range of students (Elmore, 2004). Second, research must be directed toward identifying the levels of comprehension that students achieve as a result of teacher use of instructional procedures. For example, information is needed about how students respond to questions requiring demonstration of knowledge of facts and concepts. In addition, information is needed about how students demonstrate deeper levels of comprehension that require understanding of main ideas and higher order thinking as they determine causation and make comparisons, inferences, or predictions. Finally, information is needed about how students demonstrate understanding in both written formats or objective test formats.

Instructional practices that might be combined for this purpose are questioning strategies and graphic organizers. Reports from the National Center for Education Research (Pashler et al., 2007) recommended using graphics (especially when combined with verbal descriptions) and helping students build explanations by asking and answering deep questions. These instructional components could be further enhanced if used with practices shown to benefit groups of students of diverse abilities.

Questioning Techniques

The use of questioning techniques in education is not new and has long been considered foundational to good teaching and learning, beginning with the tradition of Socrates (Elder & Paul, 1998). Indeed, it has been the subject of reviews and commentaries over recent decades (Anderson & Biddle, 1975; Andre, 1979; Dillon, 1984; Pressley et al., 1992). Increasingly, many authors have urged teachers to think about the quality of questions they use in instruction, that is, to determine what are the most critical questions for students to be able to answer.

Researchers have contended that the use of questioning strategies can guide learning (Pressley et al., 1992) and lead to increased comprehension (Craig, Sullins, Witherspoon, & Gholson, 2006; King, 1994; Palinscar & Brown, 1984; Taboada & Guthrie, 2006). Others note the importance of questioning to support problem solving and deep-level reasoning (Gholson & Craig, 2006) and to promote discourse-comprehension processes (E. Kintsch, 2005). Questioning can focus attention and help students monitor understanding (Rosenshine, Meister, & Chapman, 1996) and support the retention and transfer of information (Campbell & Mayer, 2009). In addition, some have argued that questioning techniques become even more powerful when incorporated with cognitive strategies (Rosenshine et al., 1996; Pressley, 1995).

Authors have referred to the types of questions that prompt learning in many ways. For example, Bransford et al. (2000, p. 5) described high-quality questions as "meaningful" questions, and Erickson (1998, p. 90) and Wiggins and McTighe (2005, p. 105) called them "essential questions." Lenz, Bulgren, Kissam, and Taymans (2004) used the term "critical questions" to describe the few, important, often difficult questions that all students must be able to answer in order to succeed.

However, Leinhardt cautioned that just asking a difficult question does not mean that students automatically benefit in terms of enhanced understanding (Young, 1995). In addition, asking a difficult question does not necessarily aid students in learning how to answer that question or other similar questions. Nevertheless, the careful construction and sequencing of a series of questions could conceivably provide learning scaffolds for students as they answer questions and apply knowledge to practical situations (Graesser, Person, & Hu, 2002), and engage in deep reasoning (Graesser, McNamara, & VanLehn, 2005).

It is not surprising that research on the construction of questions has shown that different types of questions elicit different types of information (Mosenthal, 1996). For example, constructing questions with interrogatives such as "why," "why not," "how," and "what if" can elicit explanations and summaries of information as opposed to facts (Graesser & Person, 1994; E. Kintsch, 2005). With regard to the sequencing of questions, recommendations have often followed educational taxonomies based on levels of thinking (Bloom, 1956). Levels of thinking are reflected in question categories (Graesser & Person, 1994) and questioning hierarchies (Taboada & Guthrie, 2006). In practice, high-level questions can facilitate deep comprehension (Rouet, Vidal-Abarca, Erboul, & Millogo, 2001) and lead students to integrate several ideas (King, 1994). Others have emphasized the use of sequences of smaller, cognitively focused and clarifying questions to enhance learning (Dantonio & Beisenherz, 2001).

As an outgrowth of this emphasis on questions, numerous studies have focused on interventions that utilize questioning techniques in some form. Rosenshine et al. (1996) reviewed the studies focusing on teaching students to generate questions as a way to improve their comprehension during or after reading or listening to a passage. In general, they found that directly teaching students to ask and answer questions yielded significant advantages on tests constructed by researchers but mixed results on standardized tests. In addition, some question types that were taught (specifically, signal words and generic questions or question stems) were more effective than others.

Similarly, Pressley et al. (1992) reviewed studies conducted when students answered questions about the content they were to learn. These authors concluded that students who are prompted with questions to generate explanatory answers, to predict the content of upcoming text, and to explain the significance of to-be-learned facts demonstrate increased learning over peers who are not prompted. Questions provided by teachers and peers are both effective in increasing student comprehension.

Graphic Organizers

Graphic organizers are two-dimensional visual aids that contain shapes and spaces into which information may be written. The organization of the shapes and spaces and the information written within them depict the relationships among the bits of key information. Graphic organizers in the form of concept maps, matrices, semantic maps, flow charts, and branching diagrams have been used to help students understand information.

Various authors have contended that the use of graphic organizers can increase students' comprehension. Reasons for these beliefs are that graphic organizers focus students' attention, activate prior knowledge, encourage active processing, and structure and organize new knowledge and relationships (Ausubel, 1968; DiCecco & Gleason, 2002; Novak, 2002; Robinson & Kiewra, 1995; Pashler et al., 2007; Wolgemuth, Trujillo, Cobb, & Alwell, 2008). The use of graphic organizers is considered particularly useful for supporting learning when combined with verbal descriptions and deep-level questioning (Pashler et al., 2007).

Research on graphic organizers as mediating devices to help students learn has a long history and has been the focus of meta-analyses of instruction that included graphic organizers (Dexter & Hughes, 2011; Moore & Readance, 1984; Nesbit & Adesope, 2006). Positive results were found for their use with students with disabilities (Bos, Anders, Filip, & Jaffe, 1989; Bulgren & Schumaker, 2006; Horton, Lovitt, & Bergerud, 1990; Kim, Vaughn, Wanzek, & Wei, 2004). Furthermore, other research has supported guidelines that facilitate essential note taking on features of critical information as cued by features on a graphic device (Boyle, 2001; Peper & Mayer, 1986).

Effective Instructional Practices

Based on Content-Enhancement principles (Bulgren, 2006), researchers at the University of Kansas Center for Research on Learning (KUCRL) have investigated the use of a combination of instructional practices called Content-Enhancement Routines (CERs). CERs exemplify effective instruction using components such as graphic organizers and specific strategies embedded within a larger set of instructional procedures (Dexter & Hughes, 2011). In CERs, *content* refers to the facts, concepts, ideas, relationships, procedures, applications, and generalizations that students must learn and use in a particular subject or discipline, such as science, social studies, mathematics or language arts. CERs are based on instructional principles designed to (a) maintain the integrity and amount of information to be delivered, (b) focus on the teacher as content expert, (c) allow the teacher to mediate learning by selecting critical features of the content and transforming them in a manner that promotes learning, and (d) teach academically diverse groups in ways that meet both group and individual needs (Bulgren, 2006; Bulgren & Lenz, 1996; Lenz, Bulgren, & Hudson, 1990; Schumaker, Deshler, & McKnight, 2002). These studies have focused on instruction using evidence-based procedures shown to help all students learn such as advance organizers and reviews, graphic organizers, reasoning strategies, and interactive co-constructed learning as teachers guide students' learning and thinking.

As shown in previous research studies, teachers can use CERs, combined with graphic organizers and strategic thinking prompts, to enhance the students' scores on tests that measure retention of factual information (Bulgren, Deshler, & Schumaker, 1997; Bulgren, Schumaker, & Deshler, 1994), comprehension of conceptual information (Bulgren, Schumaker, & Deshler, 1988), learning a concept by analogy (Bulgren et al., 2000), comparing critical concepts (Bulgren et al., 2002), and evaluating argumentation (Ellis & Bulgren, 2009). The effectiveness of these methods was tested in settings such as science, social studies, and language arts classes with students of diverse abilities, including those who were HA, AA, and LA, as well as SWDs, across a variety of content topics and a range of experimental designs.

In all of these studies, a unique graphic organizer was created to make the process of learning facts or concepts and pertinent relationships visually explicit for students. Nevertheless, although all of the studies focused on live delivery of information by a teacher, none of them addressed helping students to answer critical questions by using a scaffolded sequence of questions. Therefore, additional research on questioning is needed on the information delivered live by a teacher in classes where a population of students of diverse abilities is enrolled.

The Purpose of This Study

In this study, we examined the effects of the Question-Exploration Routine (QER), an associated graphic organizer, and an embedded strategy on students' ability to think about and answer complex questions. To do this, we focused on student achievement as measured by performance on assessments requiring students to answer questions created to test knowledge at different levels of thinking. First, we assessed student performance by analyzing the total score for all students in the study. This provided an overview of the performance of all students. Second, we assessed the performance of all students on different types of questions that assessed different levels of comprehension. Questions ranged from those that were designed to elicit knowledge of facts, concepts, or definitions to deeper levels of comprehension and higher order thinking designed to measure abilities to determine causation, make comparisons and predictions, or explain and generalize main ideas. Third, given the complexity of student populations in many classes, we explored the performance of students in different achievement groups and students with disabilities, including HA, AA, LA, and SWDs. To pursue these purposes, the experimental intervention, called the QER, was compared with a traditional lecture-discussion method.

Method

Participants

All students in the classes of a science teacher and a social studies teacher in two middle schools in a Midwestern suburban school district were invited to participate. Information on the study and consent forms were sent to the parents of all students in the classes of these teachers. The parents of 116 (out of approximately 175 possible) students returned signed consent forms allowing their children to participate. Ninety-six of the participating students were enrolled in a required seventh grade science course in one school (School A); the 20 remaining students were enrolled in a required seventh grade social studies course in another school (School B).

For purposes of assessing the efficacy of the treatment for students of different achievement and ability levels, four subgroups of students (HA, AA, LA, and SWD) were identified by analyzing participating students' grades in academic courses for each of the trimesters of the academic year during which the study was conducted. The 39 HA students (15 boys and 24 girls) had grade-point averages of 3.5 and above on a 4.0 scale. The 49 AA students (25 boys and 24 girls) had received no more than two grades below the "C" level, and the 11 LA students (10 boys and one girl) had received at least three grades below the "C" level. The 17 SWDs (11 boys and 6 girls) included students with LD and students with OHI. These students had been formally classified as such following district and state guidelines of the Kansas State Department of Education, and they had been enrolled in the participating classes as part of the district's policy promoting inclusion of SWDs.

To establish that the two groups of students were relatively the same with respect to their academic skills, their scores were compared on two standardized tests: the Iowa Test of Basic Skills–Science subtest (ITBS-Science; Hoover, Dunbar, & Frisbie, 2001) and the Cognitive Abilities Test–Verbal subtest (CAT-Verbal; Lohman & Hagen, 2001). For the ITBS-Science subtest scores, Group 1 had a mean score of 63.57 ($SD = 23.9$), and Group 2 had a mean score of 64.62 ($SD = 23.2$). These differences were not statistically significant, $F(1, 108) = .053$, $p = .82$. On the CAT-Verbal test, the scores were as follows: Group 1 $M = 58.31$ ($SD = 31.0$), and Group 2 $M = 57.23$ ($SD = 28.3$). Again, the differences were not statistically significant, $F(1, 106) = .035$, $p = .79$.

The study took place in regularly assigned classrooms in the participating schools. School A had an enrollment of approximately 560 students, with 18% of them receiving free or reduced-price lunches. The majority of students were White (90%), followed by African American (5%) and Hispanic (3%) students. School B had an enrollment of approximately 1,000 students, with 9% receiving free or reduced-price lunches. The majority of students were White (93%), followed by African American (4%) and Hispanic (2%) students.

The Intervention

The QER. We created a sequence of teaching methods called the QER to serve the purposes of this study. These methods were sequenced within three instructional phases: "Cue," "Do," and "Review." During the "Cue" phase, the teacher presents an advance organizer by (a) introducing the topic of the lesson; (b) explicitly informing students about the importance and benefits of understanding the targeted information; (c) distributing a one-page graphic organizer, called the Question-Exploration Guide (QEG), on which to take notes; and (d) prompting the students to take notes on the guide and participate in the discussion.

During the "Do" phase, the major part of the routine, the teacher and students together complete the six parts of the guide following a set of six thinking steps, prompted by the acronym ANSWER (which was constructed from the first letter of the first word of the name of each step). Specifically, the six steps are: (a) *A*sk a critical question; (b) *N*ote and explore key terms and basic knowledge needed to answer the critical question; (c) *S*earch for supporting questions and answer those supporting questions; (d) *W*ork out or formulate a clear, concise main-idea answer to the critical question; (e) *E*xplore the main-idea answer in a related area; and (f) *R*elate the main idea to today's real world. Thus, this sequence involves the explicit development of a critical question and supporting questions, along with answers to those questions, and ways that students can apply, transfer, and generalize knowledge to other course content and the world around them. These steps represent a strategy to guide students in cognitive processing related to answering questions posed in a variety of subject and grade levels.

During the instructional process, information associated with the six steps is written on the associated graphic, the QEG. (See Figure 1 for an example guide.) For example, a space is provided for posing a critical question; this space is cued by the number "1" and the question, "What is the critical question?" Then space is pro-

Question Exploration Guide

Text Reference: Chapter 5, pp. 41-44
Course / Unit / Lesson: X
Critical Question #: 2
Title: Modern Warfare: Biological Weapons
Name: David Cole
Date: May 1, 2011

① What is the Critical Question? **Why are biological weapons such a great danger?**

② What are the Key Terms and explanations?

ORGANISM - A living thing	ANTIBODY - a body's own defense against infection
BIOLOGICAL WEAPON - An organism or its poisons that cause harm	ANTIBIOTIC - a man-made substance that kills harmful things
	VACCINE - a dead organism that gives protection

③ What are the Supporting Questions and answers?

		EXAMPLES & EFFECTS	TREATMENTS & PROBLEMS
What are examples & effects of each biological weapon?	VIRUS	Smallpox causes fever/blisters;	treated with vaccine (no longer made).
	BACTERIA	Anthrax harms lungs;	treated with vaccines/antibiotics (not enough).
What are treatments & problems with treatments for each biological weapon?	FUNGUS	Wilt kills crops;	treated with fungicides (can't cover large areas).
	TOXIN	Ricin poisons & kills;	no protection (must avoid).

④ What is the main Idea answer? **Biological weapons can harm people and crops, and treatments are inadequate.**

⑤ How can we use the main idea?
What are similarities between wilt & ricin? (Both from plants.)
What are differences between wilt & ricin? (Wilt is plant; ricin is a poison from plant.)

⑥ Is there an Overall Idea? Is there a real-world use?
Why must vaccines be given before exposure to biological weapons? (Takes time to build immunity.)

Figure 1. Example question-exploration guide for the critical question, "Why are biological weapons such a danger?"

vided for key terms; this space is cued by the number "2" and the question, "What are the key terms and explanations?" Space is also provided for "unpacking" the critical question; that is, breaking it into smaller, more manageable questions, and answering them; this space is cued by the number "3" and the question, "What are the supporting questions and answers?" Then space is provided for the answer; this space is cued by the number "4" and the question, "What is the main idea answer?" Space is also provided for answering a question requiring students to apply the main idea in a related area; this space is cued by the number "5" and the question, "How can we use the main idea?" Finally, space is provided for answering a question requiring students to generalize the main idea to a real-world or transfer situation; this space is cued by the number "6" and the questions, "Is there an overall idea? Is there a real-world use?"

Finally, in the "Review" phase, the teacher and students review the information covered in the "Do" phase and the process used to answer the critical question.

The lecture-discussion method. The researchers designed this method after observing teachers in classrooms and the instructional methods they were using to teach their classes (Schumaker, Deshler, Lenz, Bulgren, & Davis, 2006). Thus, the lecture-discussion method involved the distribution of a note-taking sheet, oral delivery of information by the teacher about which students were to take notes, and the writing of items of information on an overhead projector transparency, which the students could copy onto their note-taking sheets.

The Instructional Topics and Materials

Topics. The information taught in this study focused on two topics: biological weapons and chemical weapons. These topics were selected because information related to them was similar to the type of information that students might be expected to understand. The teachers of the participating students concurred that the topics and related information would be appropriate for this age group of students and that the information had not been covered in their courses. Information was gathered for each of the topics using textbooks and recent articles on the topics.

Materials. From information gathered on these topics, researchers created four scripts: One that followed the QER sequence described above on the topic of chemical weapons, one that followed the lecture-discussion method on the topic of chemical weapons, one that followed the QER sequence on the topic of biological weapons, and one that followed the lecture-discussion method on the topic of biological weapons. Specifically, for the chemical weapons topic, there was a QER script (see Figure 1 in the online supplemental materials) and a lecture script (see Figure 2 in supplemental materials). Likewise, for the biological weapons topic, there was a QER script (see Figure 3 in supplemental materials) and a lecture script (see Figure 4 in supplemental materials).

The two scripts for each topic were parallel in several respects. For example, they contained exactly the same items of information on the same topics. The information included the same vocabulary words and the same definitions for those words, information on the topics, the development of the main idea, and use of the idea. Second, the order of information was the same based on the order of phases and steps of the QER. Third, the scripts contained the same verbal advance and postorganizer review statements. Fourth, both scripts included a place where the students were asked to discuss and help coconstruct the answer to the critical question. This was the only time when students were prompted to participate in an extensive discussion. Fifth, the time devoted to each topic was the same.

Despite these similarities, the scripts were also different in some respects. First, questions were explicitly stated in the QER script. For example, questions on the QEG related to chemical weapons were as follows: (a) "Why would a nation develop chemical weapons?" (b) "What are examples and common uses of each chemical weapon?" and (c) "What are the effects of and defenses against each type of weapon?". In contrast, the critical question was implicitly stated in the lecture-discussion scripts ("First, we will discuss the important topic of why a nation would develop chemical weapons," and "To explore information about each of these types of chemical weapons, we will talk about examples and common uses for each as well as effects of and defenses against each type of weapon. Let's look at each of these areas in order.")

For each of the QER scripts, researchers developed a QEG. (See Figure 1 for the QEG associated with biological weapons.) For each of the lecture-discussion scripts, researchers also developed a note-taking sheet similar to those observed during lecture-discussion presentations by teachers in a previous study (Schumaker et al., 2006). For the note-taking sheets, the same vocabulary words or phrases were typed on each prior to distribution to the students. For the biological weapons topic, the words were *organism*, *biological weapon*, *antibody*, *antibiotic*, *vaccine*, *virus*, *bacterium*, *fungus*, and *toxin*. For the chemical weapons topic, the words were *chemical*, *chemical weapon*, *antidote*, *atropine*, *pyridostigmine*, *tear gas*, *mustard gas*, *blood gas*, and *nerve gas*.

The materials provided for taking notes were also different in some respects. The students receiving the QER instruction received the appropriate QEG (with the critical question and nine relevant vocabulary words or phases typed on it), and the students receiving the lecture instruction received the appropriate note-taking sheet (with the nine relevant vocabulary words or phrases typed on it). During instruction, the procedures differed as follows: For the QER instruction, the teacher wrote the information shown in Figure 1 for the information on biological weapons and information in a similar format for information on chemical weapons with the appropriate QEG. For the lecture-discussion instruction, the teacher wrote the definitions for the nine vocabulary terms on the overhead transparency during the lecture-discussion instruction. (See Figure 5 in supplemental materials for the note-taking sheet for biological weapons and Figure 6 in supplemental materials for the note-taking sheet for chemical weapons.)

Three independent reviewers (two science teachers with master's degrees and a third with a bachelor's degree in science and a master's degree in reading) reviewed the parallel scripts and the sources of the information. They all indicated on a checklist that 100% of the items of information in the two pairs of parallel scripts were accurate, based on the sources. They all also indicated on the same checklist that the same items of information were represented in both scripts for each topic.

Measurement Instruments

Content test. From the information contained in the scripts, the researchers designed a test to assess student comprehension and retention of the information related to both topics. The test included multiple-choice, matching, and short-answer questions. Questions related to the two topics were intermixed randomly throughout the test. A total of 40 points could be earned on the total test: 16 matching-question points, 20 multiple-choice-question points, and four short-answer-question points. The format of the test was similar to that of assessments commonly administered in classes and allowed analysis of student responses on each type of question. The test was given on the day following the day of instruction.

The matching portion of the test contained a word bank consisting of four words representing examples of biological weapons (*fungus*, *virus*, *bacterium*, *toxin*); these words were identified with the letters "a," "b," "c," or "d." These words were placed in a column on the left side of the page. Eight descriptors representing characteristics or definitional phrases that described an example of a biological weapon ("causes smallpox" or "can be defended only by vaccines"), each of which could be correctly associated with only one of the four items in the word bank, were located in a column on the right side of the page. Next, the students were presented with another word bank consisting of four words representing examples of chemical weapons (*tear gas*, *mustard gas*, *blood gas*, *nerve gas*); these words were identified with the letters "a," "b," "c," or "d." They were also placed in a column on the left side of the page below the first word bank; eight descriptors representing characteristics or definitional phrases that described an example of a chemical weapon ("a nonliving substance" or "an antidote given after exposure to nerve gas"), each of which could be correctly associated with only one of the four items in the word bank, were also located in a column on the right side of the page.

Students were asked to associate each of the words with a descriptor by placing the letter associated with the correct word in a blank placed in front of each descriptor. This section explored factual knowledge, knowledge of the definitions of vocabulary terms, and comprehension of the characteristics of concepts. A total of 16 points was available for this portion of the test.

The multiple-choice portion of the test contained 20 questions. Sixteen questions were designed to explore levels of thinking in the form of relationships and use (e.g., comparison, causation, application, inference, or prediction). These questions assessed higher order comprehension to fulfill one of the purposes of this study. For example, we determined the ability of students to engage in the higher order thinking by asking questions that required students to contrast how wilt is different from ricin, how toxins are different from bacteria, and how biological weapons are similar in terms of causing harm. The ability to generalize use of the information was determined by questions that required understanding about and explanation of inferences or predictions, such as predicting how a country might respond to inspectors if it were found to possess sarin, inferring what officials might be expecting if the population were given atropine, or predicting the relative lasting effects of different types of weapons.

In addition, we designed two of the multiple-choice questions to assess comprehension of the main idea associated with chemical weapons and two to assess comprehension of the main idea associated with biological weapons. Each multiple-choice question was worth one point each for a possible total of 20 points on this part of the test.

Finally, on the short-answer portion of the test, students wrote short answers to show their understanding of the main ideas that were discussed and developed by the teacher and students during each lesson. For example, for the main idea associated with biological weapons, one critical component was that biological weapons can harm people and crops; the second critical component was that treatments are inadequate. For the main idea associated with chemical weapons, one critical component was that chemical weapons are made from common chemicals; the other critical component was that there are few defenses against them. Two points were awarded if both components of the main idea were written, one point was awarded if only one component was written, and zero points were awarded if neither component was written. A total of four points was possible on the short-answer portion of the test.

The same three independent researchers who determined that both scripts accurately reflected information from the source also reviewed the scripts and the test. On a checklist designed for this purpose, reviewers indicated whether each item on the test was covered in both of the parallel scripts about the given topic to which the item related. All of the reviewers indicated that 100% of the items were covered in both parallel scripts for each topic.

Implementation checklist. An observer used a specially developed checklist to determine whether the instructor covered all the components of the instruction. The checklist listed the key components of all the scripts, including (a) cueing of the topic, importance of the information, use of the graphic, and expectations for participation; (b) specifying the critical question/issue, naming the key terms and definitions of the key terms, developing information related to each topic, discussing the main idea, and extending the main idea to the subject area and to the real world; and (c) reviewing the content and process of the instruction. An observer sat in the classroom and recorded a point for each item on the checklist that was presented during each lesson.

Reliability. Two scorers independently scored a random sample of 31 (26.7%) of the students' tests for reliability purposes using a randomly sampled set of students' tests from each class in proportion to the number of students in the class. The two observers' recordings were compared item by item for an exact scoring match, and the percentage of agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. For items on the matching and multiple-choice questions, observers agreed on 100% of the items; for short-answer responses, observers' scores agreed 121 times out of 124 opportunities for 97.6% agreement.

Procedures

The study took place in the students' classrooms at the time they were scheduled to be present in that room. The first author presented all the instruction to all the students present in those classrooms on the days of the study, but data were used only for students whose parents had signed and returned consent forms. The researcher provided instruction in the five classes of the science teacher during one school day; on the next day, she provided instruction in the two participating classes of the social studies teacher.

Depending on the assignment of each class to condition, the researcher provided the specified instruction to each class within the 50-min period for 25 min per topic. The researchers distributed a pencil and the appropriate sheet on which to take notes to each student; the students were directed to take notes as specified in each script, but no specific instruction on note taking was provided. The researchers collected students' notes at the end of instruction on each topic. The researcher followed the specified lesson plan and its variation for each class by using the scripts associated with that lesson plan and writing the pertinent information on an overhead transparency.

On the day following the instruction, the researchers administered the test to all students. Students were allowed to study from their notes for 5 min at the beginning of the class period. Each student's notes were distributed at the beginning of the class period and collected by the researchers prior to the test. A researcher read the test instructions to the students, but test items were not read to the students. The researchers administered the short-answer portion of each test first and collected it prior to administration of the rest of the assessment. Students had a total of 45 min to complete the test. They were not allowed to consult their notes or each other as they took the test.

Design and Analysis

We used a counterbalanced design that involved all 116 students experiencing two instructional treatments. Seven classrooms were randomly assigned to one of two groups. Students in one group (three classes) received instruction in which the subject of chemical weapons was delivered using the QER format, and instruction on biological weapons was delivered in a lecture-discussion format. Students in another group, (four classes) received instruction on chemical weapons in a lecture format and instruction on biological weapons in the QER format. The order of the topics (chemical or biological weapons) was randomly assigned. Thus, each student received both treatments, with the order of treatment and topic randomly assigned within classes to balance order effects and control for possible group differences. In addition, we controlled for instructional time and teacher effects by having the same instructor deliver all lessons. We controlled for difficulty of information by analyzing the two topics (chemical and biological) separately.

We compared the test scores within like topics; that is, scores on the tests related to chemical weapons were compared between the groups of students receiving the instruction with the QER for chemical weapons and those who received instruction on this topic in the lecture-discussion format. Similarly, we compared the scores for biological weapons between groups of students receiving instruction for biological weapons with the QER and those receiving the lecture-discussion instruction. Statistical tests conducted to determine if there were any order effects showed no such effects for either the chemical weapons topic, $F(1,4.9) = 1.63$, $p = .26$, or the biological weapons topic, $F(1, 5.97) = 1.64$, $p = .25$.

All analyses were conducted using a hierarchical linear model (HLM) approach with SAS PROC MIXED (Version 9.1). HLM analysis is one form of general linear mixed model (GLMM) analysis. GLMM analyses are related to the familiar general linear model (GLM) approach that includes multiple regression, analysis of variance, analysis of covariance, etc. One advantage of the GLMM approach is that it allows multiple sources of variance to be modeled, whereas the GLM approach allows only one source of variance (usually the variance due to students) to be modeled. Multiple sources of variance may occur because of a nested or hierarchical data structure, such as students nested within classrooms. In the GLMM approach, both the variance among the students within the classroom and the variance between classrooms (often due to teacher effects) can be modeled. The GLMM approach results in more accurate standard errors for testing the hypotheses regarding the intervention effects in education studies. In this analysis, we used two levels: students at Level 1 nested within classrooms at Level 2. Restricted maximum likelihood estimation was used for all parameters, and the Kenard-Rogers method of determining degrees of freedom was used in all analyses.

Recent research on determining a standard effect-size measure to be used in HLM analyses has led to widespread acceptance of a value calculated using the unstandardized parameter estimate for the treatment effect divided by the raw score standard deviation. This is comparable to the Cohen *d* effect size and is interpreted the same way: .20 is small effect, .50 is medium effect, and .80 is large effect (Cohen, 1988). As noted previously, our primary purpose was to assess the effectiveness of the QER in enhancing student performance at multiple levels of thinking. We report the statistical results for the full sample on the overall test and then for each subtest used for the different levels of thinking. A secondary interest was the effectiveness of the curriculum with subgroups of students who had different achievement levels and students with disabilities. Because our total sample was somewhat small and was selected to be representative of the larger overall student population, we necessarily had some subgroups with too few students to conduct statistical tests. Therefore, for the subgroups of students, we report the results with descriptive statistics—means (in bar charts and tables), percentages, frequencies, and mean differences. We first report the results for the chemical weapons instruction and then for the biological weapons. Note that all reported test scores are percentage correct scores.

Results

Chemical Weapons Test Results

Chemical weapons overall test results. Students receiving QER instruction on chemical weapons earned significantly higher scores on the total test than students receiving the lecture-discussion format instruction, $F(1, 5.7) = 27.8$, $p = .002$. The effect size was 1.42, a very large effect. Means, number of students, and standard deviations for the total test score and the subgroup test scores for all students are presented in Table 1.

The mean percentage scores earned by students for the total group and each subgroup in both conditions are shown in Figure 2. As shown in the left half of the figure, the means indicate that students in the QER condition performed, on average, 26 percentage points better than in the traditional lecture-discussion condition. For the subgroups of students in the QER condition, all groups performed better than in the traditional lecture-discussion. The difference in performance of the students in the QER group compared to that of the lecture-discussion group was about the same for the LA, AA, and HA groups; the difference was between 28 and 30 points higher for those groups of students receiving the QER instruction. The SWDs in the QER group scored 15 points higher than in the lecture-discussion group.

Chemical weapons matching results. Students in the QER group also earned significantly higher scores than did students in the lecture-discussion group on matching questions about chemical weapons, $F(1, 108) = 23.4$, $p = <.0001$. The effect size was 1.15, a very large effect size (see the right half of Figure 3 for mean scores on matching items). Overall, students in the QER condition performed, on average, 24 percentage points better than students in the traditional lecture-discussion condition. For subgroups of students in the QER condition, all groups performed better than in the traditional lecture-discussion, and all groups had similar differences ranging from 20 to 29 percentage points. For this test, the difference between the SWDs in the QER group and the SWDs in the lecture-discussion group was 26 points; only the HA group had a greater difference.

Chemical weapons multiple-choice results. Students receiving the QER instruction also earned significantly higher scores on the multiple-choice questions than did students in the lecture-discussion format, $F(1, 5.5) = 13.7$, $p = .012$. The effect size was 1.26, a very large effect (see the left half of Figure 3 for total group

Table 1
Total Test and Subtest Scores

	Chemical weapons test scores (% correct)						Biological weapons test scores (% correct)					
	Lecture discussion			QER			Lecture discussion			QER		
Question type	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*	*M*	*N*	*SD*
Matching	50.57	66	21.64	75.00	50	22.59	53.00	50	23.22	70.08	66	23.91
Multiple choice	51.36	66	21.90	77.60	50	22.00	51.60	50	21.51	71.36	66	21.04
Main idea multiple choice	62.88	66	38.54	85.00	50	29.01	52.00	50	39.07	77.27	66	29.33
Main idea short answer	0.76	66	6.15	29.00	50	36.55	13.00	50	24.35	62.12	66	36.23
Total	45.98	66	16.92	71.70	50	18.67	48.30	50	17.54	69.92	66	19.91

Note. QER = question-exploration routine.

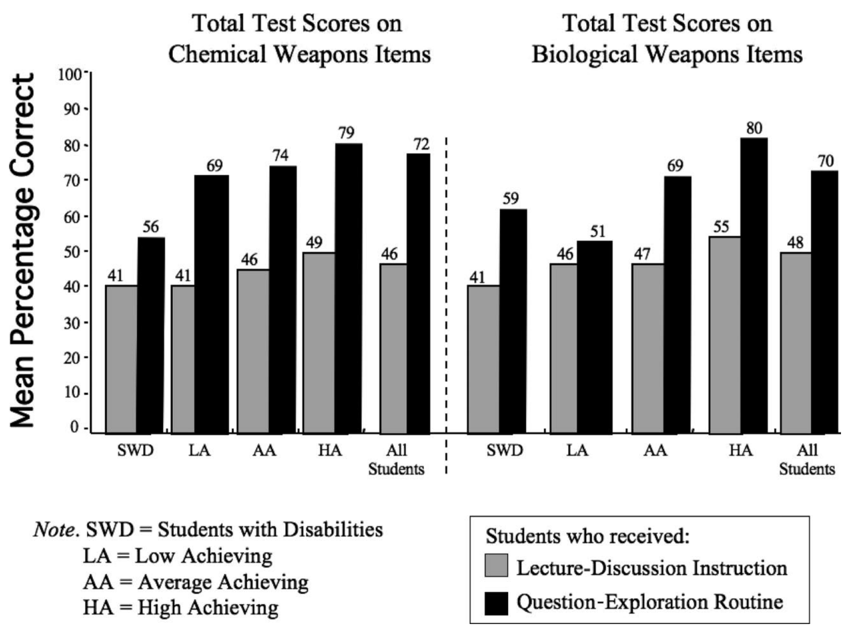

Figure 2. Mean percentage scores earned by students on total test scores. SWD = students with disabilities; LA = low achieving; AA = average achieving; HA = high achieving.

and subgroup means on multiple-choice items). As shown in Figure 3, the pattern of mean differences between the two groups on the multiple choice questions was similar to the pattern for the total test described above. For the total group of students, the difference in performance was 27 percentage points. In the subgroups, the LA, AA, and HA students had similar differences ranging from 28 to 32 percentage points. The SWDs in the QER group performed about eight points higher than the SWDs in the lecture-discussion group.

Chemical weapons main idea test results. Students were also assessed on their understanding of the main idea regarding chemical weapons. This assessment was conducted by using a

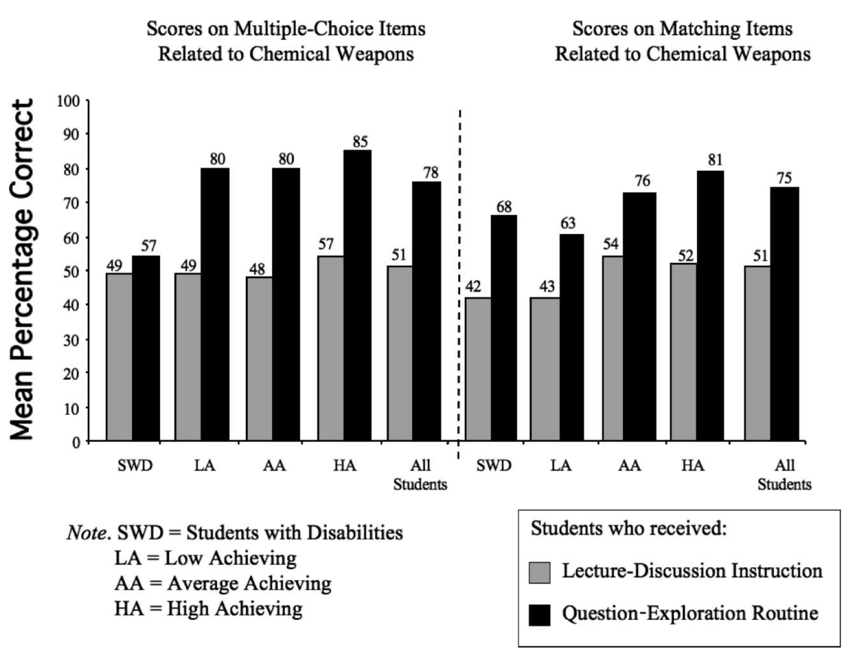

Figure 3. Mean percentage scores earned by students on multiple-choice and matching items about chemical weapons. SWD = students with disabilities; LA = low achieving; AA = average achieving; HA = high achieving.

short-answer essay format and two multiple-choice questions designed to elicit the same main idea. These scores had a limited range of 0–2 points, making the GLMM analysis questionable. We were, however, interested in comparing the patterns of the results for the main idea tests to those for the previously discussed tests, and so we conducted the GLMM analysis but supplemented it by also conducting a contingency table analysis in which participants were classified by their scores (0, 1, 2) and their experimental condition.

For the main idea multiple-choice question, the total sample QER students performed better than the lecture discussion students, $F(1,7.21) = 5.54$, $p = .05$ with an effect size of .46 (medium effect). The difference between the scores for the two groups was 22 points. In the contingency table analysis, the Pearson $\chi^2(2) = 11.3$, $p = .003$. Cohen (1988) defined a w statistic as a measure of effect size for contingency table analyses; interpretation of the w values is .10, small; .30, medium, .50 large. The w value for this analysis was .31, a medium effect. For all the subgroups, the QER group scored higher than the lecture-discussion group with differences ranging from six points (SWDs) to 36 points (the LA group). In fact, in the LA group, all students in the QER condition answered both of the questions correctly.

When using the short-answer format to test for knowledge of the main idea, the students in the QER group scored higher than those in the lecture-discussion group. The differences were not statistically significant, $F(1, 5.47) = 5.22$, $p = .067$, although the effect size was 1.46 (very large). In the contingency table analysis, Pearson $\chi^2(2) = 45.1$, $p < .0001$. The effect size w = .53, a large effect. The difference in points for the total sample was 28 points. All groups had more difficulty answering questions that required short-answer responses than those in the multiple-choice format, although the range of differences for the subgroups was similar to that for the multiple-choice main idea, from six points to 38 points. However, only the HA subgroup in the lecture-discussion group had any points on the short answer test, whereas in the QER group, every subgroup had students who earned points. The means and differences between the groups are depicted in Figure 4.

A further analysis was conducted on the chemical weapons test to determine the percentage of students who performed at a level commonly deemed as "passing" (i.e., a score of 60% or above). We report the percent passing in the lecture-discussion classes and in the QER classes for each student group. For the chemical weapons lessons, the percentages passing were as follows: SWDs, 25% in lecture-discussion, 33% in QER; LAs, 29% in lecture-discussion, 45% in QER; AAs, 26% in lecture-discussion, 82% in QER; and HAs, 33% in lecture-discussion, 93% in QER.

Biological Weapons Test Results

Biological weapons overall test results. On questions related to biological weapons, students who participated in the QER instruction earned significantly higher total scores than students who participated in the lecture instruction, $F(1, 10.2) = 18.7$, $p = .001$. The effect size was 1.16, a large effect size (see the right half of Figure 2 for mean percentage scores earned by the total group and subgroups of students). As shown on the right half of Figure 2, the means for the total test scores indicate that the students in the QER condition performed, on average, 22 points better than in the traditional lecture-discussion condition. For subgroups of students in the QER condition, all groups performed better than those in the

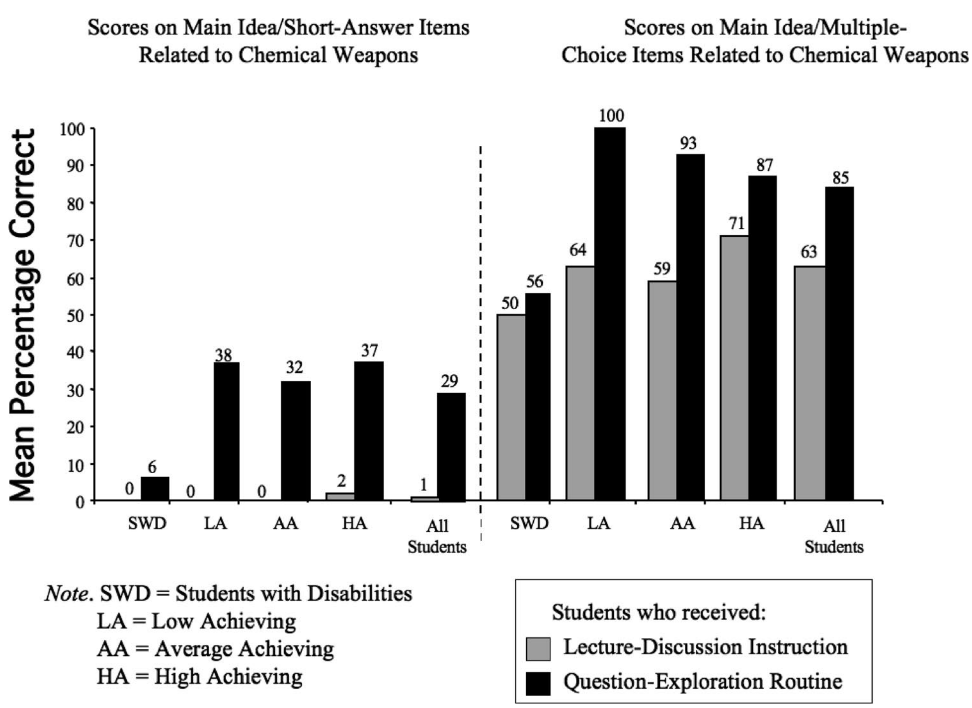

Figure 4. Mean percentage scores earned by students on short-answer and multiple-choice main-idea items about chemical weapons. SWD = students with disabilities; LA = low achieving; AA = average achieving; HA = high achieving.

traditional lecture-discussion group. In general, in the QER group, SWD, AA, and HA groups had similar differences compared with the lecture-discussion groups (18, 22, and 25 points, respectively), whereas the difference for the LA group was more limited: five points.

Biological weapons matching test results. For the matching portion of the test, the mean percentage of correct answers for students in the QER classes was significantly higher than the mean percentage scores of students taught using the lecture-discussion format, $F(1, 108) = 4.95$, $p = .028$. The effect size was .68, a medium effect size (see the right half of Figure 5 for mean percentage scores on matching items). As shown on the right half of Figure 5, the chart indicates that for the matching items on the test, students in the QER condition performed, on average, 17 percentage points better than in the traditional lecture-discussion condition. The AA and HA in the QER group did much better than the lecture-discussion group: 22 and 17 points, respectively. The differences between the two conditions were much smaller for the LA and SWD groups: zero and five points, respectively.

Biological weapons multiple-choice test results. For the multiple-choice portion of the test, students in the QER instruction groups answered a significantly higher percentage of questions correctly than the students in the lecture on biological weapons classes, $F(1, 6.85) = 12.9$, $p = .009$. The effect size was .97, a large effect size. As shown on the left half of Figure 5, the students in the QER condition performed, on average, 19 percentage points better than those in the traditional lecture-discussion condition. All subgroups of students in the QER condition performed better than those in the traditional lecture-discussion, although the difference for the LA students was only four percentage points. The biggest difference was for the SWDs, where the difference was 32 percentage points.

Biological weapons main idea test results. Students' understanding of the main idea about biological weapons was assessed in the same way as was done for the chemical weapons, including a short-answer format and two multiple-choice questions related to the main idea. For the biological weapons main idea questions, the analyses were conducted in the same way as the chemical weapons analyses, both a GLMM analysis and a contingency table analysis were performed. Figure 6 reports the mean scores for the total group and subgroup short-answer questions and the multiple-choice questions designed to assess understanding of the main idea associated with biological warfare.

When using the two multiple-choice questions designed to assess knowledge of the main idea, significant differences emerged between the two instructional groups, $F(1, 108) = 12.2$, $p = .0007$; effect size = .53, a medium effect. The contingency table analysis gave similar results, Pearson $\chi^2(2) = 15.2$, $p = .001$; effect size w = .362, a medium effect. As shown in right half of Figure 6, the means for responses to multiple-choice questions about the main idea are, on average, 25 points better for students in the QER condition than in the traditional lecture-discussion condition on the total test score. For subgroups of students in the QER condition, all groups performed better than in the traditional lecture-discussion; the range of differences was from 19 to 37 points, with the SWDs having the largest mean difference.

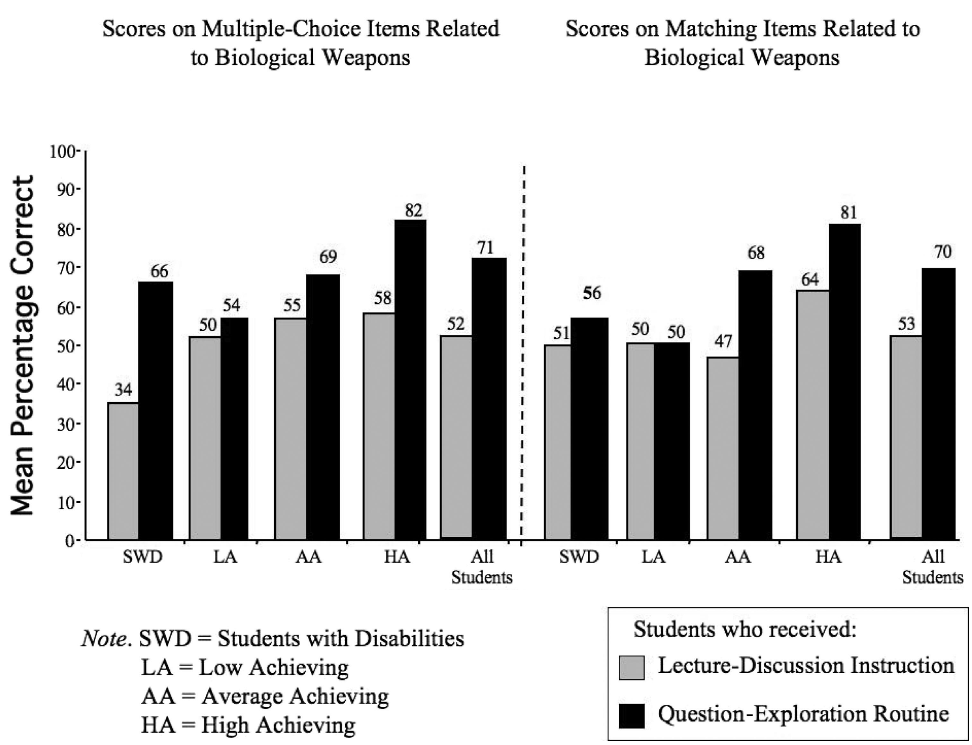

Figure 5. Mean percentage scores earned by students on multiple-choice and matching items about biological weapons. SWD = students with disabilities; LA = low achieving; AA = average achieving; HA = high achieving.

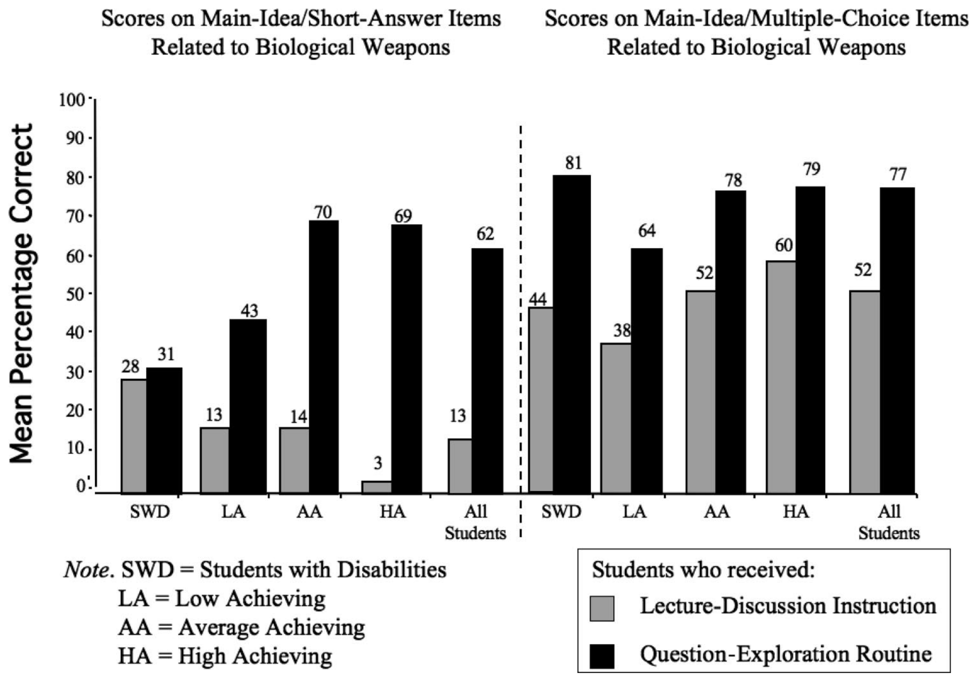

Figure 6. Mean percentage scores earned by students on short-answer and multiple-choice main-idea items about biological weapons. SWD = students with disabilities; LA = low achieving; AA = average achieving; HA = high achieving.

Using the short-answer format, significant differences were found between the two instructional groups, $F(1, 7.99) = 32.2$, $p = .0005$; effect size = 1.63, a very large effect. The contingency table analysis produced similar results, Pearson $\chi^2(2) = 45.1$, $p <$ .0001; w = .525, a large effect. As shown in the left half of Figure 6, means for the short-answer responses to questions about the main idea, students in the QER condition performed, on average, 49 percentage points better than in the traditional lecture-discussion condition. In the QER condition, all groups performed better than in the traditional lecture-discussion. Greater differences were found for students in the AA and HA groups (56 and 66 points, respectively) than in the other subgroups. It is notable that the difference for the SWDs in the QER group was only three points higher than for SWDs receiving traditional instruction.

We also determined the percentage of students who made "passing" grades on the biological weapons tests. We report the percent passing in the lecture-discussion classes and in the QER classes for each student group. For the biological weapons lessons, the percentages passing were as follows: SWDs, 33% in lecture-discussion, 50% in QER; LAs, 25% in lecture-discussion, 29% in QER; AAs, 23% in lecture-discussion, 74% in QER; and HAs, 53% in lecture discussion, 88% in QER.

Discussion

This study responded to challenges found in national standards and worldwide demands that require all students to answer meaningful, critical questions in a variety of areas. To do this, we examined the effects of a QER on students' ability to think about and answer complex questions when compared to the effects of traditional lecture-discussion instruction.

Overall, students in the QER instruction performed better than those in the lecture-discussion instruction. Statistically significant differences and very large and large effect sizes on mean total test scores were found in favor of groups of students who participated in the QER instruction on chemical weapons and biological weapons, respectively, compared with those who received the lecture-discussion instruction. Overall, very large effect sizes were found for groups of students receiving instruction about chemical weapons on two portions of the assessment: those made up of matching items that assessed knowledge of facts, vocabulary definitions, and concepts and those made of up multiple-choice questions that assessed comprehension of higher order relationships and generalization of important main ideas. For the biological weapons test, a significant difference and large effect size were also found in favor of the group receiving the QER instruction for the multiple-choice items, and a significant difference and medium effect size was found for the matching questions.

Furthermore, analysis of student responses on questions eliciting understanding of main ideas indicated that the groups receiving the QER instruction performed better on both topics than did those receiving the lecture-discussion instruction on both the short answer and multiple-choice questions. Very large effect sizes were found on both topics for short-answer responses about main ideas, and medium effect sizes were found for the multiple-choice items testing understanding of main ideas. Because students receiving the QER instruction received higher scores on every subtest except one (in which the same mean scores were found), these findings indicate that, in general, students' scores on assessments can be enhanced through the use of the QER.

The study shows that explicit oral and written presentation of a critical question, supported by logically related subquestions, can positively affect students' scores on a test partially comprised of higher order questions. It replicates the results of previous studies showing that the explicit presentation of important conceptual information combined with the use of graphic organizers and strategic thinking steps can improve the performance of groups of students representing diverse abilities. Further, because the teacher wrote factual information on an overhead transparency in both treatments, it is evident that simply modeling how to write information does not enhance outcomes (see the matching item results).

Another important issue relates to students' performance on questions requiring different types of responses, particularly written responses. On average, students in all subgroups had more difficulty answering questions that required a written response than questions in a multiple-choice format designed to elicit understanding of main ideas. Many students did not even try to answer questions that required a written response but nevertheless appeared to comprehend the main ideas when assessed in an objective format. This finding raises the question of whether the literacy goals of the No Child Left Behind Act of 2001 and the Individuals with Disabilities Improvement Act (U.S. Department of Education, 2004) associated with writing are achievable without more research on how teachers can instruct students to write better responses while teaching content.

Because the QER was designed to respond to the needs of students with a wide range of academic achievement and abilities, information was also collected and reported descriptively on the performance of subgroups of students. These subgroups included students with disabilities, students who were low achievers, average achievers, and high achievers. Upon examination, differences were found in favor of students receiving instruction utilizing the QER over students receiving instruction in a lecture format for 49 of the 50 comparisons between groups on the total score, matching items, multiple-choice items, and short-answer items. In general, these findings led to the conclusion that the QER helped students from each of these subgroups.

An important issue relates to the indications of strengths as well as the needs of diverse learners. For example, a large mean difference of 37 percentage points was found in favor of the SWDs in the group receiving instruction using the QER over SWDs in the lecture-discussion group for the questions designed to assess understanding of the main idea in a multiple-choice format related to biological weapons. The mean difference for this subgroup was larger than differences in mean scores for any of the other subgroups of students in either of the two topic groups, exceeding the overall mean score by four percentage points. However, the same large mean difference was not observed for this group of students on the multiple-choice items in the chemical weapons instruction. On average, the SWD group performed only slightly better than the students in the lecture-discussion group, achieving a score that was not in the generally passing range.

In addition, students within the LA group who received instruction using the QER, on average, outscored all other groups on the section of the test designed to assess the main idea about chemical weapons via multiple-choice items. In fact, all students who were LA and received instruction with the QER on chemical weapons answered these questions correctly, for a mean score of 100%—the only group to achieve this mean score. This was a mean difference of 36 percentage points in favor of the group of LAs who received instruction in using the QER compared with those receiving the lecture instruction. This is an important finding in view of the current emphasis on larger critical ideas in subject areas and standards. However, variability in performance was found more often for groups of LA students and SWDs than for HA and AA groups of students.

The variability in results may be due to several factors. These possible factors may include differences in prior knowledge, student interest, or inherent differences in the material. However, the variability may well be due to the small numbers of students in both the LA and SWD groups, a limitation to the study. Therefore, this supports the need for replication with larger numbers, especially for LA students and SWDs. Nevertheless, it is noteworthy that passing grades may not be easily achievable for some students without additional instruction or repeated use of a given instructional procedure to assure familiarity with the processes.

Limitations of the study include the fact that components of the QER, as originally developed, were not fully used. Because of time and design limitations, the typical amount of student discussion and group exploration of transfer and generalization of the main idea answer were not included. Furthermore, the routine was not tested in classes other than science or social studies, or when a classroom teacher rather than a researcher delivered instruction. These results suggest topics for future research.

Relative to teachers, research might focus on the fidelity of using the instructional procedures when used by classroom teachers for regularly scheduled instruction, as well as the amount of teacher time, effort, and support necessary to integrate the QER across entire units and courses. Future research might also explore teacher development and use of integrated sets of CERs, such as those designed to help students master concepts, learn by analogy, make comparisons, explain causes and effects, or evaluate options and make decisions, in addition to answering critical questions using the QER. These integrated sets might be used (a) across several content areas in the same school or district, (b) across grade levels in targeted content areas, or (c) in studies of system wide change.

Relative to students, future research needs to focus on (a) the abilities of students to acquire and generalize use of strategic approaches to thinking, such as those included in CERs, when they are used by teachers in classes primarily to teach content; (b) the usefulness of teaching the same types of higher order thinking across different subject matter (Stevens, Wineburg, Herrenkohl, & Bell, 2005); (c) the number of repeated uses of a QER necessary to achieve more positive performance outcomes for subgroups of students; (d) other supports some students may need to benefit from instruction using the QER; (e) the usefulness of CERs in enhancing discipline-specific thinking demands; and (f) instructional procedures to support students in writing responses on assessments, as required by increased writing demands (Graham, Harris, MacArthur, & Schwartz, 1991).

Relative to the mechanisms of the intervention, future research could explore the contributions of the major components in the QER, the graphic organizer and the questioning, in helping students learn. This could be accomplished by a study in which information was presented with three treatment conditions: one with the questioning techniques in the QER, one without the graphic organizer, and one with a graphic organizer only to present the information without the use of the questioning techniques. Finally, future research could explore students' cognitive processes during instruction.

In general, this study supports the use of the QER and indicates that instruction associated with answering critical questions required by complex content challenges can be incorporated into education settings for use in large-group instruction in classes containing students of diverse abilities. It contributes to ongoing research on Content-Enhancement instruction that may be integrated into whole courses to respond to several levels of learning demands: foundational knowledge and comprehension, manipulation of relationships, higher order thinking, and transfer and generalization of knowledge. Integration and use of this type of instruction could contribute to the goal of helping students engage in the type of thinking needed to succeed in today's worldwide economy.

Finally, the study underscores the important role questions can play in increasing student comprehension of targeted concepts and important ideas. Given the increased emphasis on higher order thinking in outcome assessments in schools, the QER appears to be a viable instructional procedure for promoting student knowledge and understanding. Further, with the pronounced presence of academic diversity in most classrooms, these instructional procedures could commensurately improve the outcomes of all student subgroups.

References

Anderson, R. C., & Biddle, B. W. (1975). On asking people questions about what they are reading. In G. H. Bower (Eds.), *The psychology of learning and motivation* (Vol. 9, pp. 90–131). New York, NY: Academic Press.

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research, 49,* 280–318. doi:10.3102/00346543049002280

Ausubel, D. P. (1968). *Educational psychology: A cognitive view.* New York, NY: Holt, Rinehart, and Winston.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96,* 523–535. doi:10.1037/0022-0663.96.3.523

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29,* 344–370. doi:10.1016/j.cedpsych.2003.09.002

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives.* New York, NY: Longman.

Bos, C. S., Anders, P. L., Filip, D., & Jaffe, L. E. (1989). The effects of an interactive instructional strategy for enhancing reading comprehension and content area learning for students with learning disabilities. *Journal of Learning Disabilities, 22,* 384–390. doi:10.1177/002221948902200611

Boyle, J. R. (2001). Enhancing the note-taking skills of students with mild disabilities. *Intervention in School and Clinic, 36,* 221–224. doi:10.1177/105345120103600405

Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain mind, experience and school.* Washington, DC: National Research Council Committee on Learning Research and Educational Practice.

Bulgren, J. A. (2006). Integrated content enhancement routines: Responding to the needs of adolescents with disabilities in rigorous inclusive secondary content classes. *Teaching Exceptional Children, 38,* 54–58.

Bulgren, J. A., Deshler, D. D., & Schumaker, J. B. (1997). Use of a recall enhancement routine and strategies in inclusive secondary classes. *Learning Disabilities Research & Practice, 12,* 198–208.

Bulgren, J. A., Deshler, D. D., Schumaker, J. B., & Lenz, B. K. (2000). The use and effectiveness of analogical instruction in diverse secondary content classrooms. *Journal of Educational Psychology, 92,* 426–441. doi:10.1037/0022-0663.92.3.426

Bulgren, J. A., & Lenz, B. K. (1996). Strategic instruction in the content areas. In D. D. Deshler, E. S. Ellis, & B. K. Lenz (Eds.), *Teaching adolescents with learning disabilities: Strategies and methods* (2nd ed., pp. 409–473). Denver, CO: Love.

Bulgren, J. A., Lenz, B. K., Marquis, J. G., Deshler, D. D., Schumaker, J. B., Davis, B., & Grossen, B. (2006). The instructional context of inclusive secondary general education classes: Teachers' instructional roles and practices, curricular demands, and research-based practices and standards. *Learning Disabilities: A Contemporary Journal, 4,* 39–65.

Bulgren, J. A., Lenz, B. K., Schumaker, J. B., Deshler, D. D., & Marquis, J. G. (2002). The use and effectiveness of a comparison routine in diverse secondary content classrooms. *Journal of Educational Psychology, 94,* 356–371. doi:10.1037/0022-0663.94.2.356

Bulgren, J. A., Marquis, J. G., Lenz, B. K., Schumaker, J. B., & Deshler, D. D. (2009). Effectiveness of question exploration to enhance students' written expression of content knowledge and comprehension. *Reading and Writing Quarterly, 25,* 271–289. doi:10.1080/10573560903120813

Bulgren, J. A., & Schumaker, J. B. (2006). Teaching practices that optimize curriculum access. In D. D. Deshler & J. B. Schumaker (Eds.), *Teaching adolescents with disabilities: Accessing the general education curriculum* (pp. 79–120). Thousand Oaks, CA: Corwin Press.

Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (1988). Effectiveness of a concept teaching routine in enhancing the performance of LD students in secondary-level mainstream classes. *Learning Disability Quarterly, 11,* 3–17. doi:10.2307/1511034

Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (1994). The effects of a recall enhancement routine on the test performance of secondary students with and without learning disabilities. *Learning Disabilities Research & Practice, 9,* 2–11.

Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology, 23,* 747–759. doi:10.1002/acp.1513

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conley, D. T. (2010). Rethinking college readiness. *New Directions for Higher Education, 144,* 3–13.

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level-reasoning-questioning effect: The role of dialogue and deep-level-reasoning questions during vicarius learning. *Cognition and Instruction, 24,* 565–591. doi:10.1207/s1532690xci2404_4

Dantonio, M., & Beisenherz, P. C. (2001). *Learning to question, questioning to learn: Developing effective teacher questioning practices.* Boston, MA: Allyn & Bacon.

Darling-Hammond. (2010). *The flat world and education: How America's commitment to equity will determine our future.* New York, NY: Teachers College Press.

Davey, B., & McBride, S. (1986). Generating self-questions after reading: A comprehension assist for elementary students. *Journal of Educational Research, 80,* 43–46.

Dexter, D. D., & Hughes, C. A. (2011). Graphic organizers and students with learning disabilities: A meta-analysis. *Learning Disability Quarterly, 34,* 51–72.

DiCecco, V. M., & Gleason, M. M. (2002). Using graphic organizers to attain relational knowledge from expository text. *Journal of Learning Disabilities, 35,* 306–320. doi:10.1177/00222194020350040201

Dillon, J. T. (1984). Research on questioning and discussion. *Educational Leadership, 42,* 50–56.

Elder, L., & Paul, R. (1998). The role of Socratic questioning in thinking, teaching, and learning. *Clearing House, 71,* 297–301. doi:10.1080/00098659809602729

Ellis, J. D., & Bulgren, J. (2009, January 8). *Improving teaching of scientific argumentation skills.* Paper presented at the annual meeting of the Association for Science Teacher Educators, Hartford, CT.

Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance.* Cambridge, MA: Harvard Education Press.

Erickson, L. (1998). *Concept-based curriculum and instruction: Teaching beyond the facts.* Thousand Oaks, CA: Corwin Press.

Gholson, B., & Craig, S. D. (2006). Promoting constructive activities that support vicarious learning during computer-based instruction. *Educational Psychology Review, 18,* 119–139. doi:10.1007/s10648-006-9006-3

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10,* 17–31. doi:10.1002/(SICI)1099-0720(199611)10:7<17::AID-ACP435>3.0.CO;2-7

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through point&query, autotutor, and iSTART. *Educational Psychologist, 40,* 225–234. doi:10.1207/s15326985ep4004_4

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31,* 104–137. doi:10.3102/00028312031001104

Graesser, A. C., Person, N. K., & Hu, X. (2002). Improving comprehension through discourse processing. In D. F. Halpern & M. D. Hakel (Eds.), *Applying the science of learning to university teaching and beyond: New directions for teaching and learning* (pp. 33–44). San Francisco, CA: Wiley Periodicals.

Graham, S., Harris, K. R., MacArthur, C. A., & Schwartz, S. (1991). Writing and writing instruction for students with learning disabilities: Review of a research program. *Learning Disability Quarterly, 14,* 89–114. doi:10.2307/1510517

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77,* 334–372. doi:10.3102/003465430303953

Heller, R., & Greenleaf, C. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement.* Washington, DC: Alliance for Excellent Education.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa Test of Basic Skills.* Rolling Meadows, IL: Riverside.

Horton, S. V., Lovitt, T. C., & Bergerud, D. (1990). The effectiveness of graphic organizers for three classifications of secondary students in content area classes. *Journal of Learning Disabilities, 23,* 12–22. doi:10.1177/002221949002300107

International Reading Association. (2006), *Standards for middle and high school literacy coaches.* Newark, NJ: Author.

Kim, A. H., Vaughn, S., Wanzek, J., & Wei, S. (2004). Graphic organizers and their effects on the reading comprehension of students with LD. *Journal of Learning Disabilities, 37,* 105–118. doi:10.1177/00222194040370020201

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31,* 338–368. doi:10.3102/00028312031002338

Kintsch, E. (2005). Comprehension theory as a guide for the design of thoughtful questions. *Topics in Language Disorders, 25,* 51–64. doi:10.1097/00011363-200501000-00006

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49,* 294–303. doi:10.1037/0003-066X.49.4.294

Lenz, B. K., Bulgren, J. A., & Hudson, P. (1990). Content enhancement: A model for promoting the acquisition of content by individuals with learning disabilities. In T. E. Scruggs & B. L. Y. Wong (Eds.), *Intervention research in learning disabilities* (pp. 122–165). New York, NY: Springer-Verlag.

Lenz, B. K., Bulgren, J. A., Kissam, B., & Taymans, J. (2004). SMARTER planning for academic diversity. In B. K. Lenz, D. D. Deshler, & B. Kissam (Eds.), *Teaching content to all: Evidence-based inclusive practices in middle and secondary schools* (pp. 47–77). Boston, MA: Allyn & Bacon.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test.* Rolling Meadows, IL: Riverside.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal Experimental Psychology, 55,* 51–62. doi:10.1037/h0087352

McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes, 38,* 1–30. doi:10.1207/s15326950dp3801_1

McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34,* 147–171. doi:10.2190/1RU5-HDTJ-A5C8-JVWE

Moore, D. W., & Readence, J. E. (1984). A quantitative and qualitative review of graphic organizer research. *Journal of Educational Research, 78,* 11–17.

Mosenthal, P. B. (1996). Understanding the strategies of document literacy and their conditions of use. *Journal of Educational Psychology, 88,* 314–332. doi:10.1037/0022-0663.88.2.314

National Center for History in the Schools. (1996). *National standards for history* (Basic ed.). Los Angeles, CA: Author.

National Council of Teachers of English. (1996). *Standards for the English language arts: A project of International Reading Association & National Council of Teachers of English.* Newark, DE: International Reading Association.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards.* Washington, DC: National Academy Press.

Nesbit, J. C., & Adesope, O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76,* 413–448. doi:10.3102/00346543076003413

No Child Left Behind Act of 2001, 20 U.S.C. §6301, et seq. (2007).

Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education, 86,* 548–571. doi:10.1002/sce.10032

Organization for Economic Co-Operation and Development. (2007). *PISA 2006: Science competencies for tomorrow's world, briefing note for the United States.* Paris, France: Author.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175. doi:10.1207/s1532690xci0102_1

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning: IES practice guide (NCER 2007–2004).* Jessup, MD: National Center for Education Research.

Peper, R. J., & Mayer, R. E. (1986). Generative effects of note-taking during science lectures. *Journal of Educational Psychology, 78,* 34–38. doi:10.1037/0022-0663.78.1.34

Pressley, M. (1995). *Cognition, teaching and assessment.* New York, NY: HarperCollins.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27,* 91–109. doi:10.1207/s15326985ep2701_7

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal, 30,* 523–553.

Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology, 87,* 455–467. doi:10.1037/0022-0663.87.3.455

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66,* 181–221. doi:10.2307/1170607

Rothman, R. (2009). *Common standards: The time is now.* Retrieved from http://www.all4ed.org/publication_material/TheTimeIsNow

Rouet, J. F., Vidal-Abarca, E., Erboul, A. B., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes, 31,* 163–186. doi:10.1207/S15326950DP3102_03

Schumaker, J. B., Deshler, D. D., Lenz, B. K., Bulgren, J. A., & Davis, B. (2006). *Strategies for helping adolescents with disabilities access the general education curriculum* (Final research report submitted to the U.S. Department of Education, Office of Special Education Programs). Lawrence, KS: University of Kansas Center for Research Institute for Academic Access.

Schumaker, J. B., Deshler, D. D., & McKnight, P. (2002). Ensuring success in the secondary general education curriculum through the use of teaching routines. In M. A. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 791–823). Bethesda, MD: NASP.

Stevens, R., Wineburg, S., Herrenkohl, L. R., & Bell, P. (2005). Comparative understanding of school subjects: Past, present, and future. *Review of Educational Research, 75,* 125–157. doi:10.3102/00346543075002125

Swanson, H. L., & Deshler, D. D. (2003). Instructing adolescents with learning disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities, 36,* 124–135. doi:10.1177/002221940303600205

Taboada, A., & Guthrie, J. T. (2006). Contributions of student questioning and prior knowledge to construction of knowledge from reading information text. *Journal of Literacy Research, 38,* 1–35. doi:10.1207/s15548430jlr3801_1

Torff, B. (2003). Developmental changes in teachers' use of higher order thinking and content knowledge. *Journal of Educational Psychology, 95,* 563–569. doi:10.1037/0022-0663.95.3.563

U.S. Department of Education. (2004). *IDEA'97 amendments, final regulations.* Retrieved from http://www.ed.gov/policy/speced/reg/regulations.html

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2010). *ESEA blueprint for reform: The reauthorization of elementary and secondary education act.* Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/publicationtoc.html

Vaughn, S., Gersten, R., & Chard, D. J. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children, 67,* 99–114.

Wiggins, G., & McTighe, J. (Eds.). (2005). *Understanding by design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Wolgemuth, J. R., Trujillo, E., Cobb, R. B., & Alwell, M. (2008). *The effects of visual display interventions on academic outcomes for youth with disabilities: A systematic review.* Ft. Collins, CO: Colorado State University.

Young, K. M. (1995). Explaining explanations: A conservation with NRCSL senior scientist Gaea Leinhardt. *Learning, 3,* 1–3.

Received November 23, 2009
Revision received March 14, 2011
Accepted April 4, 2011 ■