Review of Educational Research

http://rer.aera.net

Teaching Students to Generate Questions: A Review of the Intervention Studies

Barak Rosenshine, Carla Meister and Saul Chapman

REVIEW OF EDUCATIONAL RESEARCH 1996; 66; 181

DOI: 10.3102/00346543066002181

Review of Educational Research
Summer 1996, Vol. 66, No. 2, pp. 181–221

Teaching Students to Generate Questions: A Review of the Intervention Studies

Barak Rosenshine
Carla Meister
Saul Chapman
University of Illinois at Urbana

This is a review of intervention studies in which students have been taught to generate questions as a means of improving their comprehension. Overall, teaching students the cognitive strategy of generating questions about the material they had read resulted in gains in comprehension, as measured by tests given at the end of the intervention. All tests were based on new material. The overall median effect size was 0.36 (64th percentile) when standardized tests were used and 0.86 (81st percentile) when experimenter-developed comprehension tests were used. The traditional skill-based instructional approach and the reciprocal teaching approach yielded similar results.

Question generation is an important comprehension-fostering (Palincsar & Brown, 1984) and self-regulatory cognitive strategy. The act of composing questions focuses the student's attention on content. It involves concentrating on main ideas while checking to see if content is understood (Palincsar & Brown, 1984). Scardamalia and Bereiter (1985) and Garcia and Pearson (1990) suggest that question generation is one component of teaching students to carry out higher-level cognitive functions for themselves.

The first purpose of this review was to attempt to evaluate the effectiveness of this cognitive strategy. We were interested in presenting the results of those intervention studies that both taught this cognitive strategy to students and assessed whether this instruction transferred to improved student reading or listening comprehension on new materials.

Our second purpose was to use this research to help us learn how to teach cognitive strategies. Early work on teaching cognitive strategies was done by Gagne (1977) and Weinstein (1978). The teaching of cognitive strategies has also been studied by Duffy et al. (1987), Pressley et al. (1990), and Meichenbaum (1977), and we hoped to add to this previous work. To do so, we set out to identify and study the instructional procedures used in these studies. Through this study, we hoped to identify and discuss instructional concepts that might be added to our vocabulary on instruction, concepts that might be useful for the teaching of other cognitive strategies.

Cognitive Strategies

Cognitive strategies are procedures that guide students as they attempt to complete less-structured tasks such as reading comprehension and writing. The earliest use of the term appears in the work of Gagne (1977) and Weinstein (1978). We believe this concept represents a major instructional advance because it helps us focus on the value of identifying or developing procedures that students can use to independently assist them in their learning.

There are some academic tasks that can be treated as *well-structured tasks.* Such a task can be broken down into a fixed sequence of subtasks or steps that consistently lead to the same result. There are specific, predictable algorithms that can be followed to complete well-structured tasks. These algorithms enable students to obtain the same result each time they perform the algorithmic operations. These well-structured tasks have often been taught by teaching each step of the algorithm to students. The results of the research on teacher effects (Good & Brophy, 1986; Rosenshine & Stevens, 1986) are particularly relevant in helping us learn how to teach students algorithms that they can use to complete well-structured tasks.

In contrast, reading comprehension, writing, and study skills are examples of *less-structured tasks.* Such a task cannot be broken down into a fixed sequence of subtasks or steps that consistently and unfailingly lead to the desired end result. Unlike well-structured tasks, less-structured tasks are not characterized by fixed sequences of subtasks, and one cannot develop algorithms that students can use to complete these tasks. Because less-structured tasks are generally more difficult, they have also been called *higher-level tasks.* However, it is possible to make these tasks more manageable by providing students with cognitive strategies and procedures.

Prior to the late 1970s, students were seldom provided with any help in completing less-structured tasks. In a classic observational study of reading comprehension instruction, Durkin (1979) noted that of the 4,469 minutes of Grade 4 reading instruction she observed, only 20 minutes were spent in comprehension instruction by the teacher. Durkin noted that teachers spent almost all of the instructional time asking students questions, but they spent little time teaching students comprehension strategies that could be used to answer the questions.

In the late 1970s, as a result of such astonishing findings and of emerging research on cognition and information processing, investigators began to develop and validate cognitive strategies that students might be taught to help them perform less-structured tasks. In the field of reading comprehension, such strategies have included question generation and summarization (Alvermann, 1981; Paris, Cross, & Lipson, 1984; Raphael & Pearson, 1985). Cognitive strategies have also been developed and taught in mathematics problem solving (Schoenfeld, 1985), physics problem solving (Heller & Hungate, 1985; Heller & Reif, 1984; Larkin & Reif, 1976), and writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985).

A cognitive strategy is a heuristic. That is, a cognitive strategy is not a direct procedure or an algorithm to be followed precisely but rather a guide that serves to support learners as they develop internal procedures that enable them to perform higher-level operations. Generating questions about material that is read is an example of a cognitive strategy. Generating questions does not lead directly, in a step-by-step manner, to comprehension. Rather, in the process of generating questions, students need to search the text and combine information, and these processes help students comprehend what they read.

The concept of cognitive strategies represents at least two instructional advances. First, the concept itself provides us with a general approach that can be applied to the teaching of higher-order skills in the content areas. When teachers are faced with difficult areas, they can now ask, "What cognitive strategies might I develop that can help students complete these tasks?" Second, researchers have completed a large number of intervention studies in which students who were taught various cognitive strategies obtained significantly higher posttest scores than did students in the control groups. The cognitive strategies that were taught in these studies and the instructional procedures by which these cognitive strategies were taught can now be used as part of regular instruction (see Pressley et al., 1990).

Although there is an extensive knowledge base available to teachers on the teaching of well-structured tasks, there is, as yet, a limited knowledge base on how to teach less-structured tasks. Some excellent initial papers on this topic have been written (Collins, Brown, & Newman, 1990; Pressley, Johnson, Symons, McGoldrick, & Kurita, 1989), and this article is a continuation of that effort. This review focuses on the teaching of a single cognitive strategy, that of generating questions after reading or listening to a selection. Our intent in focusing on the teaching of a single cognitive strategy is to provide a clearer account of instructional issues.

Rationale for Teaching Questioning Strategies

The studies selected for this review are those in which students were taught to generate questions during or after reading or listening to a passage. This cognitive strategy has been referred to as a *comprehension-fostering* cognitive strategy (Palincsar & Brown, 1984; Collins et al., 1990). Student self-questioning is also described as a *metacognitive* or *comprehension-monitoring* activity, because students trained in question generation may also acquire heightened self-awareness of their comprehension adequacy (Palincsar & Brown, 1984; Wong, 1985).

Comprehension fostering and active processing. Students may become more involved in reading when they are posing and answering their own questions and not merely responding to questions from a teacher and/or a text. Composing questions may require students to play an active, initiating role in the learning process (Collins et al., 1990; King, 1994; Palincsar & Brown, 1984; Singer, 1978). By requiring students to inspect text, to identify main ideas, and to tie parts together, generating questions may engage them in a deeper processing of text material (Craig & Lockhart, 1972). Engaging in these active processes may lead to improved comprehension and enhanced recall of information, particularly of the central features of a passage (King, 1994).

Comprehension monitoring. Teaching students to ask questions may help them become sensitive to important points in a text and thus monitor the state of their reading comprehension (Wong, 1985). In generating and answering self-questions concerning the key points of a selection, students may find that problems of inadequate or incomplete comprehension can be identified and resolved (Palincsar & Brown, 1984). Student questioning may also aid in clarifying and setting dimensions for the hypotheses being formulated and assist in the control of premature and faulty conclusions.

Selecting Studies

All studies selected for this review provided instruction to students on how to generate questions either during or after reading a paragraph or passage. In addition, all studies that were selected contained equivalent experimental and control groups and included a transfer posttest whereby students in both groups were compared on their ability to comprehend new materials, materials that had not been used in the training.

We began our literature search by consulting the references in the critical review on question generation by Wong (1985). We then conducted a computer search through both the ERIC Silver Platter retrieval system and the *Dissertation Abstracts International* database. We also searched through programs of recent Annual Meetings of the American Educational Research Association, and we located additional references that were cited in the question generation studies we found. Whenever dissertation studies were used, we ordered and read the entire dissertations.

Included studies. A total of 26 studies met our criteria. In 17 of these studies, students were taught the single cognitive strategy of question generation. The other 9 studies involved reciprocal teaching, an instructional method in which the teacher first models the cognitive process being taught and then provides cognitive support and coaching, or scaffolding, for the students as they attempt the task (Palincsar & Brown, 1984). As the students become more proficient, the teacher fades the support. In reciprocal teaching a teacher might model the strategy of question generation after reading a paragraph and then provide support to the students as they attempt to generate questions by themselves. Reciprocal teaching is described by Collins et al. (1990) as an example of "cognitive apprenticeship," in which novices are taught the processes that experts use to handle complex tasks. In the 9 reciprocal teaching studies included in this review, students learned and practiced two or four cognitive strategies, one of which was question generation.

We included the reciprocal teaching studies because our inspection of the transcripts from those studies showed that during the reciprocal teaching dialogues at least 75% of the time was spent asking and responding to questions. We were also interested in comparing the results of the studies that taught only question generation with the results of the studies that taught two to four strategies, one of which was question generation, in the context of the reciprocal teaching dialogues. For these reasons, we included the reciprocal teaching studies but also presented separate columns of results for the reciprocal teaching studies and for the studies that used regular instruction.

Excluded studies. We excluded studies on question generation that lacked either transfer measures or true control groups. For example, we excluded studies in which the dependent measure was the students' comprehension of passages for which they had practiced generating questions during the instructional session (Andre & Anderson, 1979; Billingsley & Wildman, 1984; King, 1994; Singer & Donlan, 1982). In all of the studies we included, students were tested on their comprehension of new material.

Three studies were excluded either because they lacked equivalent control groups or because they lacked control groups altogether (Braun, Rennie, & Labercane, 1985; Gilroy & Moore, 1988; Wong, Wong, Perry, & Sawatsky, 1986). In the study by Gilroy and Moore, the experimental group students, who were below-average readers, were compared with average and above-average readers who had not received the question generation training. The other two studies, involving five and eight students, showed respectable and sometimes impressive gains for the students but did not have control groups.

Studies in which students were taught to generate questions before reading a passage (e.g., Adams, Carnine, & Gersten, 1982) were not included because generating such questions does not require inspecting text and monitoring one's comprehension. We also omitted six studies in which students were not given instruction or practice but were told simply to generate questions as they read. The studies by Frase and Schwartz (1975) and Rickards and Denner (1979) are two of the most frequently cited studies of this type. These studies tested only how well students learned the material read in the study; they did not assess whether the method improved student comprehension of new material (see Wong, 1985, on the importance of transfer tests).

Computing Effect Sizes

In order to compare the results of these studies, we computed effect sizes. For each study, this was done by calculating the difference between the means of the experimental and control groups and dividing this result by the standard deviation of the control group. When standard deviations were not available, we estimated effect sizes using procedures suggested by Glass, McGaw, and Smith (1981) and by Hedges, Shymansky, and Woodworth (1991). Estimated effect sizes are followed by the abbreviation "est." in Appendix B. We were unable to estimate effect sizes for three studies, all with nonsignificant results, in which three groups were used but standard deviations were not given. We assigned each of those three studies an effect size of zero. (We also considered assigning each of those studies the median effect size for all nonsignificant studies, which was 0.32, but chose instead the more conservative approach. In this review, the two procedures would have yielded the same results.) Average effect sizes are reported as medians rather than means because there were a number of effect sizes that were larger than 1.0, and reporting means would have inflated the results in this small sample.

An effect size expresses the difference between the mean scores of the experimental and control groups in standard deviation units. The gain (or loss) associated with an effect size can be computed by referring to a table of areas of the normal curve. An effect size of 0.36 stands for 0.36 of a standard deviation. Looking in a table of the area of a normal curve, we see that 0.36 corresponds to 0.14 of the area above the mean (above the 50th percentile). Consequently, we add 0.50 and 0.14 to arrive at 0.64. Thus, an effect size of 0.36 means that if an average student in the control group were to receive the treatment, she would now score at the 64th percentile of the control group. Similarly, an effect size of 0.87 would place this person at the 81st percentile of the control group. In the reporting of results, each effect size is followed by the corresponding percentile in which an average student would fall.

Only scores on comprehension tests that served as transfer measures were used. In some studies, both experimenter-developed comprehension tests and standardized tests were used as outcome measures. In such cases, separate effect sizes were reported for each type of outcome measure. In the analyses for which we used only a single effect size for each study, median effect sizes within those studies were reported. Results are also reported as statistically significant or not.

Three types of outcome measures were used in these studies: (a) standardized reading achievement tests, (b) experimenter-developed short-answer or multiple-choice tests, and (c) student summaries of a passage. When students were asked to summarize a passage, the ideas in the passage were usually grouped and scored according to their level of importance, and this scoring was applied to the ideas presented in the students' summaries. Passages used in experimenter-developed tests and in summarization tests ranged in length from 200 to 800 words.

Grouping the Studies

One approach for helping students learn less-structured tasks has been to provide *scaffolds*, or temporary supports (Wood, Bruner, & Ross, 1976). Scaffolds serve as aids during the initial learning of a complex skill or cognitive strategy and are gradually removed as the learner becomes more proficient. A model of the completed task and a checklist against which students can compare their work are two examples of such supports. Another type of scaffold is the *procedural prompt* (Scardamalia & Bereiter, 1985). Procedural prompts supply the student with specific procedures or suggestions that facilitate the completion of the task. Learners can temporarily rely on these hints and suggestions until they create their own internal structures (Scardamalia & Bereiter, 1985).

We believe the concept of procedural prompts is an important one that might be applied to the teaching and reviewing of other less-structured strategies, and we believe it deserves the focus we have given it in this review. A number of different procedural prompts were used in these studies to help students learn how to generate questions. As part of our instructional focus, we were interested in knowing the results that would be obtained if we first grouped the studies by the types of procedural prompts used and then compared the results obtained with the various types of prompts.

We organized the results around five different procedural prompts and included a category for three studies that did not use procedural prompts. Details on the studies, organized by procedural prompts, are presented in Appendix A. Each procedural prompt is described below. The effect sizes associated with each prompt will be presented in the Results section.

The five types of prompts were (a) signal words, (b) generic question stems and generic questions, (c) the main idea of a passage, (d) question types, and (e) story grammar categories.

It is interesting to note that although most of the studies provided rationales for teaching question generation, few investigators provided rationales for selecting specific procedural prompts.

Signal words. A well-known and frequently used procedural prompt for helping students generate questions consists of first providing students with a list of signal words for starting questions, such as *who, what, where, when, why,* and *how.* Students are taught how to use these words as prompts for generating questions. Signal words were used in 9 studies, as shown in Appendix A.

Generic question stems and generic questions. The second most frequently used procedural prompt was to provide students with generic questions or stems of generic questions. Students were given generic question stems in three studies by King (1989, 1990, 1992) and specific generic questions in the study by Weiner (1978).

Following are examples of the generic question stems used in the studies by King (1989, 1990, 1992): "How are … and … alike?" "What is the main idea of … ?" "What are the strengths and weaknesses of … ?" "How does … affect … ?" "How does … tie in with what we have learned before?" "How is … related to … ?" "What is a new example of … ?" "What conclusions can you draw about … ?" "Why is it important that … ?"

In the study by Weiner (1978), the following generic questions were provided:

1. How does this passage or chapter relate to what I already know about the topic?
2. What is the main idea of this passage or chapter?
3. What are the five important ideas that the author develops that relate to the main idea?
4. How does the author put the ideas in order?
5. What are the key vocabulary words? Do I know what they all mean?
6. What special things does the passage make me think about? (p. 5)

Main idea. In a third type of procedural prompt, students were taught or told to identify the main idea of a paragraph and then use the main idea to prompt the development of questions. Dreher and Gambrell (1985) used a booklet to teach this procedure to high school students and sixth grade students; it included the following suggestions:

(1) Identify the main idea of each paragraph.
(2) Form questions which ask for new examples of the main idea.
(3) If it is difficult to ask for a new instance, then write a question about a concept in the paragraph in a paraphrased form.

In other studies, similar procedures were taught orally to learning disabled and regular education junior high school students (Wong & Jones, 1982), to above-average junior high school students (Ritchie, 1985), to college students (Blaha, 1979), and to fourth and sixth grade students (Lonberger, 1988).

Question types. Another procedural prompt was based on the work of Raphael and Pearson (1985), who divided all questions into three types; each type is based on a particular kind of relationship between a question and its answer and the cognitive processes required to move from the former to the latter. The three types of questions are (a) a question whose answer can be found in a single sentence, (b) a question that requires integrating two or more sentences of text, and (c) a question whose answer cannot be found in the text but rather requires that readers use their schema or background knowledge. In Raphael and Pearson's study, students were taught to identify the type of question they were being asked and to decide upon the appropriate cognitive process needed to answer the question. In three studies (Dermody, 1988; Labercane & Battle, 1987; Smith, 1977), students were taught to use this classification to generate questions.

Story grammar categories. We found two studies (Nolte & Singer, 1985; Short & Ryan, 1984) in which students were taught to use a *story grammar* to help understand the narrative material they were reading. With fourth and fifth grade students, Nolte and Singer used a story grammar consisting of four elements: (a) setting, (b) main character, (c) character's goal, and (d) obstacles. Students were taught to generate questions that focused on each element. For example, for the character element they were taught that the set of possible questions included the following: "Who is the leading character?" "What action does the character initiate?" "What do you learn about the character from this action?"

No apparent procedural prompts. There were three studies that apparently did not teach any procedural prompts. No procedural prompts were mentioned in the complete dissertation of Manzo (1969), in the journal article of Helfeldt and Lalik (1976), or in the study by Simpson (1989). We wrote to Manzo and to Helfeldt, and they confirmed that no procedural prompts had been furnished in their studies. Rather, these studies involved extensive teacher modeling of questions and *reciprocal questioning*. The latter is a term used in Manzo's study to refer to turn taking between teacher and student; that is, the teacher first generated a question which the student answered, and the student then generated a question which the teacher answered. The teacher-generated questions served as models for the students. In addition, the authors of these three studies provided for extensive student practice, and this practice may have compensated for the lack of specific procedural prompts. In Manzo's study and Helfeldt and Lalik's study, the teacher led the reciprocal questioning teaching; in Simpson's study, the students practiced in pairs.

Quality of the Studies

We attempted to evaluate the quality of the instruction and design in all of the studies included in this review. We developed four criteria:

(1) Did the instructor model the asking of questions?
(2) Did the instructor guide student practice during initial learning?
(3) Was student comprehension assessed during the study?
(4) At the end of the study, was there an assessment of each student's ability to ask questions?

The first two criteria focused on the instruction in question generation. The second two criteria appeared in the work of Palincsar and Brown (1984) on reciprocal teaching, and we selected them because we thought they represented important areas of design that were related to instruction. Each study was rated on each of the four criteria, and the results will be discussed in the Results section.

Results

We grouped results separately by each type of procedural prompt. Within each prompt, we presented results separately for standardized tests and for experimenter-developed comprehension tests. The characteristics of each study are presented in Appendix A. The separate results for each study are presented in Appendix B.

We also grouped studies by the instructional approaches used. In 17 of the studies, traditional instruction was used, and the students were taught only the single strategy of question generation. In 9 of the studies, question generation was taught using the reciprocal teaching instructional approach (Palincsar & Brown, 1984). In every reciprocal teaching study, students were taught the strategy of question generation and one or three additional strategies. These additional strategies included summarization, prediction, and clarification. The specific strategies taught in each study are listed in Appendix A. Because of the interest in reciprocal teaching, we kept the two groups of studies separate in most of our analyses.

Table 1 presents the overall results for the 26 studies, grouped by type of test and by instructional approach. There were no differences in effect sizes between those studies that used a traditional, teacher-led approach and those that used the reciprocal teaching approach. For both approaches, the effect sizes and the number of significant results were fairly small when standardized tests were the outcome measure. For standardized tests, the median effect size was 0.36. Again, an effect size of 0.36 suggests that a person at the 50th percentile of the control group would be at the 64th percentile had she received the treatment. The effect size was 0.87 for the 16 studies that used experimenter-developed comprehension tests and 0.85 for the 5 studies that used a summarization test. An effect size of 0.87 suggests that a person at the 50th percentile of the control group would be at the 81st percentile had she received the treatment. This pattern, favoring experimenter-developed comprehension tests over standardized tests, was found across both instructional approaches.

Overall, then, the practice of teaching students to generate questions while they read has yielded large and substantial effect sizes when experimenter-developed comprehension tests and summarization tests were used. Because there was no practical difference in overall results between the two types of experimenter-developed comprehension tests—short-answer tests and summary tests—we combined the results for these two types of tests in subsequent analyses. Much smaller effect sizes were obtained when standardized tests were used.

TABLE 1
Overall effect sizes by type of test

	Instructional approach		
Type of test	Reciprocal teaching ($n = 9$)	Regular instruction ($n = 17$)	Combined ($n = 26$)
Standardized	0.34 (6)	0.35 (7)	0.36 (13)
Exp. short answer	1.00 (5)	0.88 (11)	0.87 (16)
Summary	0.85 (3)	0.81 (2)	0.85 (5)

Note. n = number of studies. Number in parentheses refers to the number of studies used to compute an effect size.

Analysis by Quality of Instruction

As noted previously, we evaluated the quality of instruction and design in each study. The results of this analysis are presented in Appendix C. All of the studies met the first two criteria: modeling and guiding practice. Therefore, we did not believe that any of the studies should be classified as low in quality. We further grouped studies according to whether they included (a) assessment of both student comprehension and student ability to generate questions, (b) assessment of either student comprehension or student ability to generate questions, or (c) assessment of neither student comprehension nor student ability to generate questions. We grouped studies into these three categories and then looked at effect sizes separately for standardized tests and separately for experimenter-developed comprehension tests. No differences were found in effect sizes among the groupings. Therefore, all studies were retained, and we continued our analyses.

Analyses by Procedural Prompts

In addition to the analysis shown in Table 1, we grouped the studies according to the different procedural prompts used to help students learn self-questioning. We did this grouping in order to explore whether using different procedural prompts would yield different results. Three of the prompts were used in at least four studies: signal words, generic questions or question stems, and main idea. Only two or three studies used each of the remaining procedural prompts, question types and story grammar. The results, by procedural prompts, are summarized in Table 2. Additional information on the individual studies, classified under each procedural prompt, is presented in Appendix A.

Signal words. Only one of the six studies that provided students with signal words such as *who* and *where* obtained significant results when standardized tests were used (median effect size = 0.36). All seven studies that used experimenter-developed comprehension tests obtained significant results. The overall median

TABLE 2
Overall median effect sizes by type of prompt

	Standardized test			Experimenter-developed test		
Prompt	Reciprocal teaching ($n = 6$)	Regular instruction ($n = 7$)	Combined ($n = 13$)	Reciprocal teaching ($n = 7$)	Regular instruction ($n = 12$)	Combined ($n = 19$)
Signal words	0.34 (4)	0.46 (2)	0.36 (6)	0.88 (5)	0.67 (2)	0.85 (7)
Generic questions/ stems					1.12 (4)	1.12 (4)
Main idea		0.70 (1)	0.70 (1)	1.24 (1)	0.13 (4)	0.25 (5)
Question type	0.02 (2)	0.00 (1)	0.00 (3)	3.37 (1)		3.37 (1)
Story grammar					1.08 (2)	1.08 (2)
No facilitator		0.14 (3)	0.14 (3)			

Note. n = number of studies. Number in parentheses refers to the number of studies used to compute an effect size.

effect size when experimenter-developed comprehension tests were used was a substantial 0.85 (80th percentile). This procedural prompt was used successfully in Grades 3 through 8.

Generic questions and generic question stems. Investigators obtained strong, significant results in almost all studies that provided students with generic questions or question stems. An overall effect size of 1.12 (87th percentile) was obtained for the four studies that used experimenter-developed comprehension tests (King, 1989, 1990, 1992; Weiner, 1978). All results on the experimenter-developed comprehension tests were significant except in the study by Weiner, in which only one of two treatment groups was significantly superior to the control group. This procedural prompt was used successfully with students at grade levels ranging from sixth grade to college, but we found no studies using this prompt in lower grades.

Main idea. In five studies students were instructed to begin the questioning strategy by finding the main idea of a passage and using it to help develop questions. Two of these studies obtained significant results for one of the two ability groups in each study (Blaha, 1979; Lonberger, 1988) The effect size was 0.70 (76th percentile) for the single study that used a standardized test and 0.25 (60th percentile) for the five studies that used experimenter-developed comprehension tests.

Question types. In studies where students were taught to develop questions based on the concept of text-explicit, text-implicit, and schema-based questions, results were nonsignificant in all three cases where standardized tests were used to assess student achievement. Only one study (Dermody, 1988) used experimenter-developed comprehension tests, and the effect size of 3.37 in Table 2 is based on that single study. However, Dermody taught both question generation and summarization through reciprocal teaching. She then assessed students on their ability to summarize new material. It is not possible to say that learning question generation alone led to Dermody's significant results.

Story grammar categories. Two studies taught students to begin with a story grammar and use the story grammar as a prompt for generating questions about the narratives they read (Nolte & Singer, 1985; Short & Ryan, 1984). Story grammar questions included questions about setting and about main characters, their goals, and obstacles encountered on the way to achieving or not achieving those goals. Both studies obtained significant results and produced an average effect size of 1.08 (86th percentile). In both studies, the teacher first modeled the questions and then supervised the students as they worked in groups asking each other questions.

No procedural prompt. Three studies did not use any procedural prompt, and all three of these used standardized tests. Helfeldt and Lalik (1976) found that the students in the experimental group were superior to control students on a post-experiment standardized reading achievement test. No differences were found between the two groups in the Manzo (1969) study or in the Simpson (1989) study. The median effect size for the three studies, all of which used standardized tests, was 0.14 (56th percentile). Both Manzo (1969) and Helfeldt and Lalik (1976) used reciprocal questioning to teach students to generate questions. None of the three studies used experimenter-developed comprehension tests.

Summary. Overall, teaching students the strategy of generating questions has yielded an effect size of 0.36 (64th percentile), compared to control group students, when standardized tests were used to assess student comprehension. When experimenter-developed comprehension tests were used, the overall effect size, favoring the intervention, was a median of 0.86 (81st percentile). There was no difference in effect size between studies that used reciprocal teaching and those that taught question generation using the more traditional format. Because the differences in results for the two types of tests were large, we did not combine the two types for an overall effect size, and we present separate effect sizes for the two types of tests throughout this review.

Which prompts were most effective? When standardized tests were used, and when we consider only those prompts that were used in three or more studies, then signal words was the most effective prompt (median effect size = 0.36; 64th percentile). Results were notably lower when the question type prompt or no prompt was used. When experimenter-developed comprehension tests were used—and again we base this result on prompts that were used in three or more studies—then generic questions or question stems and signal words were the most effective prompts. The four studies that used generic questions or question stems had an overall effect size of 1.12 (89th percentile), and the five studies that used signal words had an overall effect size of 0.85 (80th percentile).

In summary, regular instruction and reciprocal teaching yielded similar results. For both approaches, effect sizes were much higher when experimenter-developed comprehension tests were used. The most successful prompts were signal words and generic question stems.

Results by Settings

Grade level of students. The grade levels of the students in these studies ranged from third grade through college, as shown in Table 3. Overall, we found both significant and nonsignificant results in all grades from Grade 3 to Grade 9. Only one of the four studies in which third grade students were taught obtained significant results; however, the experimenters in the other three studies used standardized tests as outcome measures, and this may be the reason for the nonsignificant results in these studies. As shown in Table 1, the effect sizes in all grades were much lower when standardized tests were used. Significant results were obtained across all five studies in which college students were taught to generate questions.

Length of training. The median length of training for studies that used each type of procedural prompt is shown in Table 4. We uncovered no relationship between length of training and significance of results. The training period ranged from 4 to 25 sessions for studies with significant results, and from 8 to 50 sessions for studies with nonsignificant results.

TABLE 3
Grade level of student

Prompt	Significant	Mixed	Nonsignificant
Signal words	3 6 7 7–8	4, 7 5–6 7	3 3
Question types		4	3 5
Main idea	4, 6	6	6
	College	6, 8, 9	
Generic questions or question stems	9 College College	6	
Story grammar	4 4–5		
No facilitator	5		7–25 year olds 6

Instructional group size. The median instructional group sizes for the different types of procedural prompts are presented in Table 5. There were no apparent differences in the numbers of students in studies that had significant, mixed, and nonsignificant results. Within the studies with significant results, the number of students in each group ranged from 2 to 25; within the studies with mixed results, the range was from 3 to 22; and within the studies with nonsignificant results, the number of students in a group ranged from 1 to 25.

Type of student. The type of student receiving instruction in each study is listed in Table 6. One might classify these students into three groupings: (a) average and above-average students, (b) students who were near grade level in decoding but poor in comprehension, and (c) students who were below average in both decoding and comprehension. This third group includes students labeled in the studies as "poor readers," "learning disabled," "below average," and "remedial." Both

TABLE 4
Median length of training (in numbers of sessions)

Prompt	Significant	Mixed	Nonsignificant
Signal words	13	13	29
Question types		24	21
Main idea	17	2	10
Generic questions or question stems	7	18	
Story grammar	7		
No facilitator	14		20

TABLE 5
Median instructional group size

Prompt	Significant	Significant/ nonsignificant	Nonsignificant
Signal words	18	5	13
Question types		6.5	15
Main idea	18	17	1 (computer)
Generic questions or question stems	18	17	
Story grammar	17		
No facilitator	2.5		1

significant and nonsignificant results were obtained in studies that employed students in the first and third categories, that is, average and above-average students and students who were below average in both decoding and comprehension. Both significant and mixed results were obtained for students in the second category, that is, students who were near grade level in decoding but poor in comprehension.

One might predict that these procedural prompts would be more effective with below-average students, who need them most, and least successful with the above-average students, who are already engaging in comprehension-fostering activities. In two studies, below-average students did make greater gains than did other students in the same studies (MacGregor, 1988; Wong & Jones, 1982). However, in six studies above-average students made significantly greater gains than did comparable control students. In three of these studies, college students were taught to generate questions (Blaha, 1979; King, 1989, 1992); in another (King, 1990), above-average high school students were the participants. In the study by Dermody (1988), average students had posttest scores that were significantly

TABLE 6
Type of student

Prompt	Significant	Mixed	Nonsignificant
Signal words	Average	Below Average	Ave. and Above
	All	Good/Poor	All
	Good/Poor	Ave. and Above	
	Good/Poor		
Question types		All	LD
			All
Main idea	All	Normal(ns)/LD(s)	All
	All		All
Generic questions or question stems	Honor Students	All	
	All		
	Remedial		
Story grammar	All		
	Poor Readers		
No facilitator	Ave.		Remedial
			Remedial

superior to those of comparable control students, whereas the poor students in the study did not have posttest scores superior to those of similar control students. Overall, the results in these studies do not support the belief that below-average students benefit more from these interventions than do above-average students.

Type of instructional approach. Reciprocal teaching was used in 9 studies. In these studies, students learned a combination of two or four cognitive strategies, one of which was question generation. The remaining 17 studies taught students question generation by mean of a more traditional instructional approach. Table 1 compares the effects of these two instructional approaches according to the three different outcome measures: (a) standardized tests, (b) experimenter-developed multiple-choice or short-answer tests, and (c) tests in which students were asked to summarize a passage and the results were scored either for the level of the propositions or total propositions. Overall, there was no difference in scores between regular teaching and reciprocal teaching when the comparison was based on standardized tests. The reciprocal teaching studies were slightly favored when experimenter-developed short-answer tests were used, and there was no difference when summarization tests were used.

We also attempted to compare reciprocal teaching and regular instruction studies that used the same procedural prompt, as shown in Table 2. Unfortunately, when the 26 studies were distributed across the two instructional approaches, two types of outcome measures (i.e., standardized tests and experimenter-developed comprehension tests), and six procedural prompts, the number of studies in each cell was too small to merit comparison. We did, however, compare results for those studies that used the most frequently used procedural prompt, signal words. When the procedural prompt was signal words and experimenter-developed comprehension tests were used, the effect size was 0.88 (81st percentile) for the five reciprocal teaching studies and 0.67 (75th percentile) for the two studies that used traditional instruction. The four studies in which the generic question stems and generic questions prompt was used yielded an effect size of 1.12 (87th percentile), and none of these was a reciprocal teaching study.

Based on these data, it is difficult to say whether the reciprocal teaching procedure, with its attendant instruction in two or four cognitive strategies, yields results that are superior to traditional instruction in the single strategy of question generation. Both approaches appear viable and useful.

Discussion

Summary of Results

Overall, teaching students to generate questions on the text they have read resulted in gains in comprehension, as measured by tests given at the end of the intervention. All tests contained new material. The median effect size was 0.36 (64th percentile) when standardized tests were used, and 0.86 (81st percentile) when experimenter-developed comprehension tests were used. The traditional skill-based instructional approach and the reciprocal teaching approach yielded similar results.

When procedural prompts were characterized by type and analyzed separately, signal words and generic question stems obtained the highest effect sizes. When we analyzed results according to the grade level of the students being taught, the length of training, the instructional group size, and the type of student receiving the intervention instruction, we found no differences among these subgroups.

Traditional Teaching and Reciprocal Teaching

Most of the studies used traditional procedures to teach the single strategy of question generation. That is, these studies included some form of teacher presentation, teacher modeling, and teacher guidance of student practice. A number of studies included question generation as one of two or four cognitive strategies that were taught and practiced within the context of reciprocal teaching.

In this review, Tables 1 and 2 allow us to compare the results of studies involving traditional instruction and reciprocal teaching. In general, the median effect sizes were very similar for the two approaches. Table 1 compares the two approaches on three different types of assessments: standardized tests, experimenter-developed short-answer tests, and tests that asked students to summarize a passage. Results for the two instructional approaches were very similar across the three types of tests. A comparison of the two approaches by types of prompt (Table 2) did not show any clear differences.

Thus, the traditional approach that taught only one cognitive strategy and the reciprocal teaching approach that taught four cognitive strategies obtained similar results. It would be interesting to explore these findings in future studies and determine what effect, if any, additional strategies are providing.

Standardized Tests and Experimenter-Developed Comprehension Tests

When investigators used experimenter-developed comprehension tests, the median effect size was fairly large, and the results were usually significant regardless of the type of procedural prompt used. But effect sizes were small and results were seldom significant when standardized tests were used, regardless of the type of procedural prompt and the instructional approach used. Results were significant in 16 of the 19 studies that used experimenter-developed comprehension tests, but results were significant in only 3 of the 13 studies that used standardized tests. This pattern, favoring results from experimenter-developed comprehension tests over those from standardized tests, also appears in Slavin's (1987) review on the effects of mastery learning. Slavin reported a median effect size of 0.04 when standardized tests were used, and a median effect size of 0.27 when experimenter-developed comprehension tests were used.

This phenomenon is more sharply illustrated when one looks at Appendix B and notes the results obtained in the six studies that used both standardized tests and experimenter-developed comprehension tests (Blaha, 1979; Brady, 1990; Cohen, 1983; Dermody, 1988; Lysynchuk, Pressley, & Vye, 1990; Taylor & Frye, 1992). When experimenter-developed comprehension tests were used, significant results were obtained in all six studies, with a median effect size of 0.86 (81st percentile). However, when standardized tests were administered, significant results were obtained in only two of the same six studies, with a median effect size of 0.46 (68th percentile).

One explanation for this difference may lie in the type of text material used in these studies. The practice text used in these studies, as well as the text material used in the experimenter-developed comprehension tests, appeared more "considerate" (Armbruster, 1984); the paragraphs tended to be organized in a main-idea-and-supporting-detail format with explicit main ideas. In contrast, the paragraphs and passages that we inspected in the standardized tests did not have such a clear text structure.

The differences in results obtained with standardized tests and experimenter-developed comprehension tests might also be attributed to differences in what is required to answer questions on the two types of tests. Many of the questions on standardized tests appeared to require additional, background knowledge. Examples included questions that asked why some words were italicized, questions that required additional vocabulary knowledge, questions that required inference beyond the text, questions that asked why a story had been written, and questions that asked where a passage might have appeared.

Theory and Practice

The authors of these studies often provided theory to justify their research on question generation. They noted that question generation was a means of providing active processing, central focusing, and other comprehension-fostering and comprehension-monitoring activities. However, none of the authors provided any theory to justify using specific procedural prompts. The theoretical rationale for studying question generation does not provide a teacher or an investigator with specific information on how to develop prompts or how to teach question generation. As a result of this gap between theory and practice, investigators developed or selected a variety of different prompts. These different prompts included signal words, generic question stems and generic questions, the main idea of a passage, question types, and story grammar categories. Although investigators started with the same theory, they developed prompts that were different from each other both in form and in apparent effectiveness. These differences suggest that the theory is more metaphorical than practical. The theory of active processing and comprehension-fostering activities remains important, but it does not suggest the particular pedagogical devices that can be used to help teach this processing.

Similarly, the instructional theory of scaffolding, or supporting the learner during instruction, does not suggest the specific pedagogical devices that can be used to provide scaffolding in practice. We believe this same problem exists for the teaching of other cognitive strategies such as summarization, writing, and mathematical problem solving. There is a gap between theory and practice, and the instructional inventions used to bridge that gap do not flow from the theory.

Characteristics of Successful Procedural Prompts

Procedural prompts are strategy-specific scaffolds that investigators have found or developed to help students learn cognitive strategies. Of course, the procedural prompts used in studies in this review were designed to help students learn to generate questions, but procedural prompts are not limited to studies of question generation. They have been used successfully to help students learn cognitive strategies in other content areas, including writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985) and problem solving in physics (Heller & Reif, 1984; Heller & Hungate, 1985). Pressley et al. (1990) have compiled a summary of research on instruction in cognitive strategies in reading, writing, mathematics, vocabulary, and science, and in almost all of the studies summarized by them student learning was mediated by the use of procedural prompts.

At the present time, developing procedural prompts appears to be an art. Procedural prompts are specific to the strategies they are designed to support, and because of this specificity it is difficult to derive any prescriptions on how to develop effective procedural prompts for cognitive strategies in reading, writing, and subject matter domains. Future research might focus on improving our understanding of procedural prompts and how they work, and such understanding might provide suggestions for the future development of procedural prompts. Toward that end, we now describe (a) an analysis of the characteristics of successful and less successful prompts and (b) an experimental study that explicitly compared two prompts.

Why were some procedural prompts more successful than others? Overall, the three most successful prompts were (a) signal words, whereby students were provided with words such as *who, why, how,* and *what* for generating questions; (b) generic question stems (e.g., "Another example of … was …") and generic questions (e.g., "What details develop the main idea?") that could be applied to a passage; and (c) story grammar categories, which helped students generate questions focusing on four elements of a story—the setting, the main character, the main character's goal, and the obstacles encountered by the main character.

These more successful prompts—signal words, generic questions or question stems, and story grammar categories—all appeared fairly easy to use. These prompts served to guide and focus the students but did not demand strong cognitive skills.

Studies in which students were taught to first find the main idea and then use this main idea as a source for generating questions were not as successful as studies that used signal words or generic questions or question stems. It may be that developing a main idea is a more complex and difficult cognitive task, and this may explain why the main idea prompt is less effective.

Studies in which students were taught to generate questions based on question types were notably unsuccessful when standardized tests were used. In the studies on which the notion of question types is based (Raphael & Pearson, 1985; Raphael & Wonnacott, 1985), students were taught to identify and recognize three types of questions: factual, inferential, and background-based. In those studies it was demonstrated that such recognition helped students select the cognitive processes appropriate for answering different types of questions on comprehension tests. In contrast, in the studies included in this review, students were taught to use this classification to develop their own questions. Perhaps the question types prompt is not as effective for learning to ask questions.

In summary, we speculate that the three more successful prompts provided more direction, were more concrete, and were easier to teach and to apply than the two less successful prompts. However, the effects of some of the less successful prompts might be improved through extensive instruction. We found one study that provided excellent and extensive training in using a question prompt to summarize a passage (Wong et al.,1986). Although this study was not counted in the meta-analysis because it lacked a control group, we were very impressed by the instructional procedures that were used in it. In this study, adolescents who were learning disabled and underachieving were taught a summarizing strategy.

The instruction lasted 2 months and continued until the students achieved at least 90% accuracy across 3 successive days. During the instruction, students were provided with a self-questioning prompt that included the following four points and questions:

(1) What's the most important sentence in this paragraph? Let me underline it.
(2) Let me summarize the paragraph. To summarize I rewrite the main idea sentence and add important details.
(3) Let me review my summary statements for the whole subsection.
(4) Do my summary statements link up with one another?

Students wrote summaries of passages throughout the training. Before the training began, less than 20% of the summaries written by the students were scored as correct. At the end of the training, all the students' summaries were scored as 70% correct or higher. Such gain with adolescents who have learning disabilities is impressive and may have occurred because the investigators continued instruction until mastery was achieved. This example illustrates that a seemingly difficult-to-learn strategy might be more successful in the future if instruction, practice, and feedback are continued until mastery is achieved.

Comparison between prompts. Another approach toward understanding how prompts achieve their effects would be to conduct studies in which different groups of students were taught to use different prompts and the results were compared. King and Rosenshine (1993) conducted one such study, in which they compared the effects of two different procedural prompts, the signal words prompt and the generic question stems prompt.

In King and Rosenshine's (1993) study, one group of average-achieving fifth grade students received training in the use of generic question stems (e.g., "Explain how ...") to help generate questions on a lesson they heard. Another group received training in the use of signal words (e.g., *what, why*) to generate questions on the lesson, and a third group was not given any prompt but still practiced asking questions about the lesson. All students in the three training groups did their practice in dyads.

The transfer test consisted of a new lesson, followed by discussion in pairs using the prompts, followed by testing on the material in the new lesson. Total scores for students who received and practiced with generic question stems were significantly superior to scores for students in the control group (effect size = 1.25, 89th percentile). Total scores for students who received the signal words were also significantly superior to scores for control students (effect size = 0.41, 66th percentile). When separate analyses were made for scores on just the inferential questions on the exam, the effect sizes were 2.28 (99th percentile) for those receiving the generic question stems and 1.08 (86th percentile) for those receiving the signal words prompt.

Thus, students who received either of the two procedural prompts had total posttest scores that were significantly superior to those of the control group, who received no prompt. The differences were largest for the inferential questions on the tests. When the generic question stems and signal words prompts were compared to each other, students who received and practiced with the generic question stems had posttest scores that were superior to those of the students who received the signal words (effect size = 0.48, 68th percentile). These differences were even larger when the results for inferential questions were analyzed separately (effect size = 0.60, 73rd percentile).

We cite this experimental study to illustrate the value of comparing different procedural prompts. Because the students in this study worked in pairs when studying for the transfer test, and because the students did not work independently, we thought it wise not to include this study in the present review. However, including this study would have had little effect on the median effect sizes.

The value of generic questions. The King and Rosenshine (1993) study suggests that the generic question stems provided more help for the students than did the signal words prompt, and these results were particularly strong when scores on inferential questions were studied. This finding is supported by Anderson and Roit (1993), who, as part of a larger treatment, developed a group of "thought-provoking, content-free questions" that low-scoring adolescents were taught to use in their groups as they read and discussed expository passages. The following are some of the questions used: "What do you find difficult?" "How can you try to figure that out?" "What did you learn from this?" "What are you trying to do?" Students in the intervention group scored significantly higher than control students on standardized tests taken at the end of the semester.

Generic questions and question stems appear to allow students to ask deeper, more comprehensive questions than they could have developed using signal words such as *where* and *how*. Stems such as "How does … affect …?," "What does … mean?," "What is a new example of …?," and "Describe …" may have been more effective because they promote deeper processing, initiate recall of background knowledge, require integration of prior knowledge, and provide more direction for processing than might be obtained through the use of the more simplified signal words.

Summary on successful prompts. Signal words, generic questions and question stems, and story grammar categories were the more successful procedural prompts in this review. When two of these more effective prompts, signal words and question stems, were compared with each other, the question stems prompt yielded superior results. We speculate that, in this case, the more successful prompts were more concrete, provided more direction, and allowed students to ask deeper, more comprehensive questions. The main idea prompt, an apparently more complex procedural prompt, was not as effective, possibly because students needed more instruction before they could use it successfully.

Limitations of Prompts?

We would like to explore two issues on the use of prompts: (a) the possibility that some prompts may "overprompt" and (b) the distinction between providing prompts for students and encouraging students to generate their own prompts.

Overprompting. One possible limitation of prompts is that they may overprompt. That is, they may provide so much support that the student does little processing and, as a result, does not develop internal structures. The development of internal structures is critical because reading comprehension, like all less-structured tasks, cannot be reduced to an algorithm. Students must complete most of the task of comprehension by themselves and must develop the internal structures that enable them to do so. The most prompts can do is to serve as heuristics, or general guides, as students learn to approach the task.

Although overprompting is a potential problem, it was not apparent in any of the studies in this review. Even in studies where students were given generic question stems or were provided with generic questions, the students who were given those prompts obtained significantly higher posttest comprehension scores on new material than did those students who were not given these prompts.

Knudson (1988) attempted to study overprompting in writing. In one treatment (Treatment 1), students received a typical procedural prompt. They were given five story elements: the main character, the main character's enemy, the setting, the plot, and the conclusion. Students were then instructed to write five paragraphs, each paragraph containing at least five sentences about one of the elements, but received no further prompting. In a second treatment (Treatment 2), students were given more explicit ideas for composing the five sentences in each of their paragraphs about the elements. For example, the students were told, "What does the main character look like? Describe that person." Knudson found that Treatment 1, which contained the more typical procedural prompt, produced superior writing. Treatment 2 produced more mechanical, fill-in-the-blank responses. Knudson's Treatment 2, then, was a case where more prompting was less successful than less prompting.

One can argue that the prompt in Knudson's (1988) Treatment 2 was overly prescriptive, that it prompted fill-in-the-blank behavior. However, King's (1989, 1990, 1992) generic question stems "How does ... affect ...?" and "What is a new example of ...?" could also be superficially labeled as fill-in-the-blank prompts, and yet the effect sizes in King's studies were quite high. Similarly, Nolte and Singer (1985), who also obtained high effect sizes, provided very explicit questions such as "What action does the character initiate?" In these cases, even with the support of these general questions, there was still a good deal of cognitive work a student had to do to complete the task. Nonetheless, it would seem worthwhile to identify and study specific situations in which specific prompts did not help students.

Generating versus providing prompts. Another alternative to providing prompts is to encourage students to develop their own prompts and strategies. This is an interesting idea; unfortunately, however, we did not find studies in which students in the treatment groups (or the control groups) were asked to develop their own prompts. There were 3 studies for which the investigators told us, by letter and phone, that no prompts had been provided (Helfeldt & Lalik, 1976; Manzo, 1969; Simpson, 1989). In those studies, teachers and students took turns asking questions without discussing prompts. These 3 studies, all of which used standardized tests, yielded a median effect size of 0.14 (56th percentile). There were 10 additional studies in which standardized tests were also used and specific prompts were taught (see Appendix B). The median effect size for these 10 studies was 0.36 (64th percentile). These differences are not substantial, and the numbers are small, but in this limited case providing and teaching prompts yielded higher effect sizes than not providing prompts.

In the studies in which students are taught to summarize, they are almost always provided with prompts (e.g., Armbruster, Anderson, & Ostertag, 1987; Baumann, 1984; Hare & Borchart, 1985). Thus, it will be of interest to study whether the distinctions made here between generating questions and providing questions will also appear when we inspect procedural prompts developed for teaching other cognitive strategies.

A Review of the Instructional Elements in These Studies

The previous section described the different procedural prompts used to help teach question generation and compared the effectiveness of these prompts in improving reading comprehension. This section attempts to identify and discuss other instructional elements that were used in these studies to teach the cognitive strategy of question generation. These elements might add to our knowledge of instruction, expand our teaching vocabulary, and provide direction for the teaching of other cognitive strategies.

We located these instructional elements by studying the procedures section of each study and abstracting the specific elements used during the instruction. We identified nine major instructional elements that appeared in these studies:

(1) Provide procedural prompts specific to the strategy being taught.
(2) Provide models of appropriate responses.
(3) Anticipate potential difficulties.
(4) Regulate the difficulty of the material.
(5) Provide a cue card.
(6) Guide student practice.
(7) Provide feedback and corrections.
(8) Provide and teach a checklist.
(9) Assess student mastery.

Although no single study used all nine instructional elements, all of these elements were used in different studies and in different combinations to help teach the cognitive strategy of question generation.

The validity of these elements cannot be determined by this review alone but rather will have to be determined by (a) testing these elements in experimental studies and (b) determining whether these elements appear in studies that teach other cognitive strategies.

Scaffolding

Many of these instructional elements, to be described in the following paragraphs, might be organized around the concept of scaffolding (Palincsar & Brown, 1984; Wood et al., 1976). A scaffold is a temporary support used to assist a student during initial learning. Scaffolding refers to the instructional support provided by a teacher to help students bridge the gap between current abilities and a goal. This instructional support may include prompts, suggestions, thinking aloud by the teacher, guidance as students work through problems, models of finished work that allow students to compare their work with that of an expert, and checklists that a student can use to develop a critical eye for their own work (Collins et al., 1990; Palincsar & Brown, 1984). Scaffolding makes sense for the teaching of cognitive strategies precisely because they are strategies and not step-by-step instructions for approaching the specific manifestation of any less-structured task.

Although many of the scaffolds described below did not appear in the earlier teacher effects literature (see Good & Brophy, 1986), these scaffolds seem compatible with that literature and seem applicable to the teaching of a wide range of skills and strategies. The nine forms of scaffolding and other instructional elements we identified in the studies in this review are described and discussed in the following paragraphs.

Provide Procedural Prompts

One new instructional feature not found in the teacher effects research is the use of strategy-specific procedural prompts such as generic question stems. These prompts served as scaffolds for the teaching of the strategies. Of the 23 studies on question generation, all but 3 taught procedural prompts. As noted earlier, prompts have been used to assist student learning in writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985), physics (Heller & Hungate, 1985; Heller & Reif, 1984; Larkin & Reif, 1976), and mathematical problem solving (Schoenfeld, 1985).

Provide Models of Appropriate Responses

Modeling is particularly important when teaching strategies such as question generation for completing less-structured tasks because we cannot specify all the steps involved in completing such tasks. Almost all of the researchers in these studies provided models of how to use the procedural prompts to help generate questions. Models and/or modeling were used at three different points in these studies: (a) during initial instruction, before students practiced, (b) during practice, and (c) after practice. Each approach is discussed here.

Models during initial instruction. In seven of the studies, the teachers modeled questions based on the procedural prompts. Thus, Nolte and Singer (1985) modeled questions based on elements of the story grammar (e.g., What action does the leading character initiate? What do you learn about the character from this action?). In studies that used instructional booklets (Andre & Anderson, 1979; Dreher & Gambrell, 1985) or computers to present the material (MacGregor, 1988), students received models of questions based on the main idea and then practiced generating questions on their own.

Models given during practice. Models were also provided during practice. Such modeling is part of reciprocal teaching (Palincsar, 1987; Palincsar & Brown, 1984). In reciprocal teaching, the teacher first models asking a question and the students answer. Then, the teacher guides students as they develop their own questions to be answered by their classmates, and the teacher provides additional models when the students have difficulty. Other studies also provided models during practice (Braun et al., 1985; Gilroy & Moore, 1988; Helfeldt & Lalik, 1976; Labercane & Battle, 1987; Manzo, 1969).

Another form of modeling, thinking aloud by the teacher while solving problems, also appeared in these studies. Garcia and Pearson (1990) refer to this process as the teacher "sharing the reading secrets" by making them overt. Thinking aloud is also an important part of a cognitive apprenticeship model (Collins et al., 1990).

Models were also used as a form of corrective feedback during practice. For example, in one study a computer was used to mediate instruction, and students could ask the computer to provide an example of an appropriate question if their attempt was judged incorrect (MacGregor, 1988). Simply using models, however, may not guarantee success. In a review of methods for teaching writing, Hillocks (1987) found that merely presenting great literature as models was not an effective means of improving writing.

Models given after practice. Three studies provided models of questions for the students to view after they had written questions relevant to a paragraph or passage (Andre & Anderson, 1979; Dreher & Gambrell, 1985; MacGregor, 1988). In these studies, the instruction was delivered by such means as instructional booklets or computers. The intent of the model was evaluative, to enable the students to compare their efforts with that of an expert (Collins et al., 1990).

Anticipate Potential Difficulties

Another instructional scaffold found in these question generation studies was anticipating the difficulties a student is likely to face. In some studies, the instructor anticipated common errors that students might make and spent time discussing these errors before the students made them. For example, in the study by Palincsar (1987), the teacher anticipated the inappropriate questions that students might generate. The students read a paragraph followed by three questions one might ask about the paragraph. The students were asked to look at each example and decide whether or not that question was about the most important information in the paragraph. One question could not be answered by the information provided in the paragraph, and the students discussed why it was a poor question. Another question was too narrow—it focused only on a small detail—and the students discussed why it also was a poor question. The students continued through the exercise, discussing whether each question was too narrow, too broad, or appropriate.

Another example of anticipating problems can be found in the study by Cohen (1983), where students were taught specific rules to discriminate (a) a question from a nonquestion and (b) a good question from a poor one: A good question starts with a question word. A good question can be answered by the story. A good question asks about an important detail of the story.

Although only two studies (Cohen, 1983; Palincsar, 1987) discussed this scaffold of anticipating student difficulties, this technique seems potentially useful and might be used for teaching other skills, strategies, and subject areas.

Regulate the Difficulty of the Material

Some of the investigators began by having students begin with simpler material and then gradually move to more complex materials. For example, when Palincsar (1987) taught students to generate questions, the teacher first modeled how to generate questions about a single sentence. This was followed by class practice. Next, the teacher modeled and provided practice on asking questions after reading a paragraph. Finally, the teacher modeled and then the class practiced generating questions after reading an entire passage.

Similarly, in studies by Andre and Anderson (1979) and Dreher and Gambrell

(1985), students began with a single paragraph, then moved to a double paragraph, and then moved to a 450-word passage. Another example comes from the study by Wong et al. (1986), in which students began by generating questions about a single, simple paragraph. When the students were successful at that task, they moved to single, complex paragraphs and lastly to 800-word selections from social studies texts.

In another study (Wong & Jones, 1982) the researchers regulated the difficulty of the task by decreasing the prompts. First, students worked with a paragraph using procedural prompts. After they were successful at that level, they moved to a longer passage with prompts and finally to a passage without prompts.

Provide a Cue Card

Another scaffold found across these studies was the provision of a cue card containing the procedural prompt. This scaffold seems to support the student during initial learning, as it reduces the strain upon the working memory. With a cue card, students can put more effort into the application of a strategy without using short-term memory to store the procedural prompts. For example, Billingsley and Wildman (1984) provided students with cue cards listing the signal words (e.g., *who, what, why*) that could be used as prompts for generating questions. Singer and Donlan (1982) presented a chart listing the five elements of a story grammar that the students were taught to use as prompts for generating questions. Furthermore, Wong and Jones (1982) and Wong et al. (1986) gave each student a cue card that listed the steps involved in developing a question about a main idea. In all four of these studies, the investigators modeled the use of the cue card.

Cue cards were also used in studies where students were provided with generic questions. In these studies (Blaha, 1979; Wong et al., 1986) students were provided with cue cards listing specific questions to ask after they had read paragraphs and passages (e.g., "What's the most important sentence in this paragraph?"). King (1989, 1990, 1992) provided students with cue cards showing question stems (e.g., "How are ... and ... alike?," "What is a new example of ...?").

Guide Student Practice

Some form of guided practice occurred in all of the studies we examined. Three types of guided practice are (a) teacher-led practice, (b) reciprocal teaching, and (c) practice in small groups.

Teacher-led practice. In many of the studies, the teacher provided guidance during the students' initial practice. Typically, the teacher guided students as they worked through text by giving hints, reminders of the prompts, reminders of what was overlooked, and suggestions of how something could be improved (Cohen, 1983; Palincsar, 1987; Wong et al., 1986). This guided practice was often combined with the initial presentation of the strategy, as in the study by Blaha (1979), where the teacher first taught a part of a strategy, then guided student practice in identifying and applying the strategy, then taught the next part of the strategy, and then guided student practice. This type of guided practice is the same as the guided practice that emerged from the teacher effects studies (Rosenshine & Stevens, 1986).

Reciprocal teaching. Reciprocal teaching was another form of guided practice. As noted earlier, in reciprocal teaching the teacher first models the cognitive process being taught and then provides cognitive support and coaching (scaffolding) for the students as they attempt the task. As the students become more proficient, the teacher fades the support and students provide support for each other. Reciprocal teaching is a way of modifying the guided practice so that students take a more active role, eventually assuming the role of coteacher. This approach was particularly suited to the task of learning to generate questions, because teacher and student could take turns asking and answering questions.

Practice in small groups. In some studies students met in small groups of two to six without the teacher; practiced asking, revising, and correcting questions; and provided support and feedback to each other (King, 1989, 1990, 1992; Nolte & Singer, 1985; Singer & Donlan, 1982). Such groupings allow for more support when students are revising questions and for more practice than can be obtained in a whole-class setting. Nolte and Singer applied the concept of diminishing support to the organization of groups. In their study, students first spent 3 days working in groups of five or six, then spent 3 days working in pairs, and eventually worked alone.

Provide Feedback and Corrections

Providing feedback and corrections to the students most likely occurred in all studies, but was explicitly mentioned in only a few. We identified three sources of feedback and corrections: the teacher, other students, and a computer.

Teacher feedback and corrections occurred during the dialogues and guided practice as students attempted to generate questions. Feedback typically took the form of hints, questions, and suggestions. Group feedback was illustrated in the three studies by King (1989, 1990, 1992) and in the study by Ritchie (1985). In the King studies, after students had written their questions, they met in groups, posed questions to each other, and compared questions within each group. The third type of feedback, computer-based feedback, occurred in the computer-based instructional format designed by MacGregor (1988). In this study, students asked the computer to provide a model of an appropriate question when they made an error.

Provide and Teach a Checklist

In some of the studies, students were taught self-checking procedures. For example, in the study by Davey and McBride (1986), a self-evaluation checklist was introduced in the fourth of five instructional sessions. The checklist listed the following questions: How well did I identify important information? How well did I link information together? How well could I answer my questions? Did my "think questions" use different language from the text? Did I use good signal words?

Wong and Jones (1982) wrote that students in their study were furnished with the "criteria for a good question," although these criteria were not described in the report. In the three studies by King (1989, 1990, 1992) students were taught to ask themselves the question "What do I still not understand?" after they had generated and answered their questions.

Checklists were introduced into lessons at different times in the different studies. Wong and Jones (1982) and King (1989, 1990, 1992) presented checklists during the presentation of a strategy, whereas Davey and McBride (1986) presented them during the guided practice, and Ritchie (1985) presented them after initial practice.

Assess Student Mastery

After guided practice and independent practice, some of the studies assessed whether students had achieved a mastery level and provided for additional instruction when necessary. On the fifth and final day of instruction in their study, Davey and McBride (1986) required students to generate three acceptable questions for each of three passages. Smith (1977) stated that student questions at the end of a story were compared to model questions, and reteaching took place when necessary. Wong et al. (1986) required that students achieve mastery in applying self-questioning steps, and students had to continue doing the exercises (sometimes daily for 2 months) until they achieved mastery. Unfortunately, the other studies cited in this review did not report the level of mastery students achieved in generating questions.

A Comparison With Research on Effective Teaching

How do the instructional procedures in these studies that taught cognitive strategies compare with the instructional procedures in the earlier research on teacher effects (Good & Brophy, 1986; Rosenshine & Stevens, 1986), where the focus was on the teaching of explicit skills? The two areas of research showed a number of common instructional procedures. For example, both areas contained variables such as presenting material in small steps, guiding initial student practice, providing feedback and corrections, and providing for extensive independent practice.

However, six interesting instructional scaffolds that were found in the question generation research did not appear in the teacher effects literature. The use of procedural prompts never appeared in the teacher effects literature. Another instructional procedure, the use of models, has long been in the psychological literature (see Bandura, 1969; Meichenbaum, 1977), but the concept of a teacher modeling the use of the procedure being taught did not appear in the teacher effectiveness literature. Anticipating and discussing areas where students are likely to have difficulties, regulating the difficulty of the material, providing cue cards, and providing students with checklists for self-evaluation are also instructional procedures that appeared in the question generation literature but did not appear in the extensive review of the teacher effects literature by Good and Brophy (1986). This second set of instructional procedures does not conflict with those instructional procedures that were developed in the earlier teacher effects literature. That is, we believe the instructional variables that emerged in this review are also applicable to the teaching of explicit skills.

When we compare the instructional procedures from the earlier teacher effects literature with those that emerged from this cognitive strategy research, we find that these two sets of instructional procedures complement each other. The contributions of the teacher effects research focused on the concepts of review, guided practice, diminishing prompts, and providing for consolidation. Its deficiency, however, was in not providing the scaffolds which appear important for teaching higher-order cognitive skills. The cognitive strategy studies, on the other hand, have contributed the concepts of procedural prompts and scaffolds but were less explicit on review and practice variables. Both sets of concepts are important for instruction.

Suggestions for Future Research

Several topics for future research on instruction emerged from this review and our discussion of the results.

Research on procedural prompts. Procedural prompts have been developed and taught in other areas of study, such as writing and science. One topic for study might be to compare the effectiveness of different prompts that have been used in a specific domain. Another topic would be to attempt to identify features associated with successful prompts and to test these hypotheses in intervention studies.

In this review, no evidence was found to support the suggestion that providing students with procedural prompts would limit their comprehension. However, this issue does deserve further study. It would seem worthwhile to study whether there are some prompts or types of prompts that inhibit or limit the learning of cognitive strategies. Such analyses should be done by subject area. We might also attempt to identify situations in which prompts may not be helpful.

In this review, we organized the results around the procedural prompts that were furnished to students. We believe that such an approach may be useful when reviewing other areas of research on cognitive strategy instruction, and we hope that this idea can be tested in future reviews.

Providing versus generating prompts. The studies in this review strongly suggest the value of providing students with prompts. Students who were provided with prompts made considerably greater gains in comprehension—on experimenter-developed comprehension tests—than did students in control groups. In addition, studies in which students were provided with prompts yielded higher effect sizes than did studies in which students practiced generating questions but were not provided with prompts. Nonetheless, it would seem worthwhile to continue this exploration and conduct studies comparing the effect of providing prompts to students with that of asking students to generate their own prompts or their own cognitive strategies. For example, a study on question generation might include three treatments: (a) providing students with prompts, (b) telling students to practice the strategies without giving them any explicit facilitation, and (c) a no-instruction control group. This design might be used for other reading comprehension strategies such as summarization as well as for cognitive strategy research in writing and study skills. In such studies, it is important that the treatments with prompts contain those prompts that were most effective in prior research. Perhaps through studies such as these we can add to the discussion about generating versus providing prompts.

Reducing the complexity of the task. One interesting instructional procedure, but one that appeared in only a few studies, is that of initially reducing the complexity of the strategy. A common approach to reducing the complexity was to begin the practice by generating questions about a single sentence or a single paragraph and then gradually increasing the amount of text the students have to process. It would be interesting to study the effects of simplification alone, as a single component, or in combination with other instructional elements.

Including more difficult material. The difficulty of the material to be comprehended was not explored in these studies. The type of practice material used in all or most of these studies, as well as the material used in the experimenter-developed comprehension tests, was "considerate" (Armbruster, 1984); that is, the paragraphs tended to be organized in a main-idea-and-supporting-detail format with a very explicit main idea. In contrast, the paragraphs and passages in the standardized tests that we inspected did not have such a clear format. We wonder what results might be obtained if students were to begin their study with the more user-friendly passages but then practice, discuss, and receive support while using the more difficult, complex, less-considerate passages. Perhaps such an approach might lead to improved scores on the standardized tests. The value of prompts may be greatly affected by the difficulty level of the material. We believe it is worthwhile to explore how well the different prompts would work with more difficult material.

Studying effects on students of different ages and abilities. There were too few studies to permit a discussion of the interaction between age and type of prompt. Sixth grade was the lowest grade to receive the question stem prompt. We wonder how well this prompt would work with younger and/or less able students. Another topic for study would be whether older, more able students would benefit more from an abstract prompt such as question types than they would from a concrete prompt such as signal words. Unfortunately, less than one fourth of the studies performed separate analyses for students of different abilities. We hope future researchers will design their studies so they can conduct these analyses.

Studying the use of checklists. Five of the studies in this review included checklists, but the use of checklists and the effects of different types of checklists have not been studied. It would be useful to conduct experimental studies in which the use of a checklist is contrasted with the absence of a checklist, and in which specific and more general checklists are compared for students at different ability levels.

Studying the effect of variations in the length of training. We did not find that the length of training was associated with program success. The total amount of training and practice ranged from 2 hours to 12 hours, and no apparent pattern was found. Within the five successful studies that used the signal word procedural prompt, the training and practice time ranged from 2.5 hours to 12 hours. One way to study how much time is needed would be monitor student acquisition of the skill and continue training until mastery is achieved. This monitoring occurred in Wong and Jones (1982), where instruction continued until students achieved an 80% level of mastery, but this procedure was not found in the other question generation studies.

Recommendations for Practice

Based on these results, we recommend that the skill of question generation be taught in classrooms. However, we would recommend, at present, that only two procedural prompts be used: (a) signal words and (b) generic questions or question stems. The median effect sizes for the two prompts were 0.85 (80th percentile) and 1.12 (89th percentile), respectively. The data also suggest that students at all skill levels would benefit from being taught these strategies.

Although procedural prompts have been useful in reading and other content areas, one must be aware that even well designed procedural prompts cannot replace the need for background knowledge on the topic being studied. Procedural prompts are most useful when the student has sufficient background knowledge and can understand the concepts in the material. Procedural prompts and the use of scaffolds cannot overcome the limitations imposed by a student's insufficient background knowledge.

Summary and Conclusions

The first purpose of this review was to summarize the research on teaching students to generate questions as a means of improving their reading comprehension. A second purpose was to study whether applying the concept of procedural prompts can help increase our understanding of effective instruction. To accomplish this second purpose, we organized the review around the strategy-specific procedural prompts that were provided to help students develop facility in generating questions. Different prompts yielded different results in these studies, and so grouping intervention studies by procedural prompt and then comparing results seemed a more productive strategy than simply combining all studies into a single effect size. We suggest that future reviews of studies in other areas of cognitive strategy research, such as writing and summarization, be organized around the different procedural prompts used in those studies. Such an approach might be useful for increasing our understanding of why specific studies were successful or unsuccessful.

The most successful prompts for facilitating the reading of expository text when experimenter-developed comprehension tests on expository material were used were (a) signal words and (b) generic questions or question stems. Story grammar was also successful in the two studies where it was used, but both of these studies used narrative text. We speculate that in this case these three prompts were easiest for the students to use. An apparently more complex procedural prompt, using the main idea as a prompt to generate questions, was not as effective, possibly because students need more instruction before they can use this prompt. However, these comments are speculative, and as suggested earlier we encourage more research on procedural prompts. Such research might include comparing different types of prompts and analyzing the components of successful prompts so that we might learn to develop and apply new prompts for use in instruction.

A third purpose of this review was to identify and discuss some of the instructional elements that were used to teach the cognitive strategy of question generation. This review has revealed a number of instructional elements, or scaffolds (Palincsar & Brown, 1984; Wood et al., 1976), that served to support student learning. These scaffolds include using procedural prompts or facilitators, beginning with a simplified version of the task, providing modeling and thinking aloud, anticipating student difficulties, regulating the difficulty of the material, providing cue cards, and using a checklist. These scaffolds provide us with both a technical vocabulary and tools for improving instruction. We do not know how many of these scaffolds can be applied to the teaching of writing or to the teaching of problem solving in math, physics, or chemistry. Future research that focuses on teaching cognitive strategies in other content areas may extend our understanding.

APPENDIX A

Studies that taught question generation

Study	Length (in sessions)	Strategies taught	Group size	Grade level(s)	Type of student	Delayed or follow-up assessment
			Signal words prompt			
Brady, 1990 (RT)	25	Predicting, clarifying, questioning, summarizing	6	7, 8	Below average (Native Alaskans)	2 weeks delayed and 3 months delayed
Cohen, 1983	6	Questioning	24	3	Below 85% on question generating pretest	No
Davey & McBride, 1986	5	Questioning	24	6	All	No
Lysynchuk et al., 1990 (RT)	13	Predicting, clarifying, questioning, summarizing	3–4	4, 7	Poor comp./ good decoders	No
MacGregor, 1988	12	Questioning	12	3	Average and good readers	No
Palincsar, 1987 (RT)	25	Predicting, clarifying, questioning, summarizing	11.5	Middle school	Poor comp./ good decoders	5 days
Palincsar & Brown, 1984 (RT)	20	Predicting, clarifying, questioning, summarizing	2	7	Poor comp./ good decoders	1 month
Taylor & Frye, 1988 (RT)	11	Questioning	22.5	5, 6	Average and above	No
Williamson 1989 (RT)	50	Predicting, clarifying, questioning, summarizing	14	3	All	2 weeks (over Christmas vacation)

APPENDIX A *(continued)*

Study	Length (in sessions)	Strategies taught	Group size	Grade level(s)	Type of student	Delayed or follow-up assessment
			Generic questions or question stems prompt			
King, 1989	4	Questioning	9	College	All	No
King, 1990	7	Questioning	15	9	Honor students	10 day retention of strategy use
King, 1992	8	Questioning	19	College	Remedial reading and study skills course	1 week retention of content
Weiner, 1977	18	Questioning	8	6	All	No
			Main idea prompt			
Blaha, 1979	14	Questioning	25	College freshmen	All	No
Dreher & Gambrell, 1985	2	Questioning	17	6	All (boys only)	4 days, 9 days (retention of content), 18 days (proced. and condit. knowledge of strategy)
Lonberger, 1988 (RT)	20	Questioning, summarizing	18	4, 6	All	No
Ritchie, 1985	MISQ: 18, SQ: 9	Questioning, main idea	1	6	All	No
Wong & Jones, 1982	2	Questioning, main idea	15	6, 8–9	Normal and LD	No

APPENDIX A *(continued)*

Study	Length (in sessions)	Strategies taught	Group size	Grade level(s)	Type of student	Delayed or follow-up assessment
			Question types prompt			
Dermody, 1988 (RT)	24	Predicting, clarifying, questioning, summarizing	5–8	4	All	No
Labercane & Battle, 1987 (RT)	28	Predicting, clarifying, questioning, summarizing	3–5	5	Learning disabled	No
Smith, 1977	13	Questioning	25	3	All	No
			Story grammar prompt			
Nolte & Singer, 1985	10	Questioning	19	4, 5	All	No
Short & Ryan, 1984	3	Questioning	14	4	Poor readers	3 days, 6 days (generalization test in use of strategy)
			No prompt			
Helfeldt & Lalik, 1976	14	Questioning	3–4	5	Average	No
Manzo, 1969	30	Questioning	1	7–25 years old	Remedial (summer tutorial)	No
Simpson, 1989	10	Questioning	13	6	4 or more months below grade level in reading	No

Note. A study marked "(RT)" used reciprocal teaching as an instructional approach.

APPENDIX B

Question generation studies: Dependent measures and effect sizes

<table>
<tr><th></th><th></th><th colspan="2">Experimenter-developed test</th><th colspan="2">Summary test</th></tr>
<tr><th>Study</th><th>Standardized test</th><th>Multiple-choice</th><th>Short-answer</th><th>Levels</th><th>Total</th></tr>
<tr><td colspan="6">Signal words prompt</td></tr>
<tr><td>Brady, 1990 (RT)</td><td>0.36[a]</td><td></td><td>0.87*</td><td></td><td></td></tr>
<tr><td>Cohen, 1983</td><td>0.57*[b]</td><td></td><td></td><td>0.57* (est)</td><td></td></tr>
<tr><td>Davey & McBride, 1986</td><td></td><td></td><td>Literal 0.62*
Infer. 0.91*
Average 0.77</td><td></td><td></td></tr>
<tr><td>Lysynchuk et al., 1990 (RT)</td><td>0.55[a,c]</td><td></td><td>0.83*</td><td></td><td>0.52* (s/ns)</td></tr>
<tr><td>MacGregor, 1988</td><td>0.35[d]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Palincsar, 1987 (RT)</td><td></td><td></td><td>1.08* (est)</td><td>0.68* (est)</td><td></td></tr>
<tr><td>Palincsar & Brown, 1984 (RT)</td><td></td><td></td><td>1.0* (est)</td><td></td><td></td></tr>
<tr><td>Taylor & Frye, 1988 (RT)</td><td>0.07[a]</td><td></td><td></td><td>0.85*</td><td></td></tr>
<tr><td>Williamson, 1989 (RT)</td><td>0.32[e]</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Generic questions or question stems prompt</td></tr>
<tr><td>King, 1989</td><td></td><td colspan="2">1.37* (est)</td><td></td><td></td></tr>
<tr><td>King, 1990</td><td></td><td colspan="2">1.70*</td><td></td><td></td></tr>
<tr><td>King, 1992</td><td></td><td>0.87*</td><td></td><td></td><td></td></tr>
<tr><td>Weiner, 1977</td><td></td><td>0.78*</td><td>0.48</td><td></td><td></td></tr>
<tr><td></td><td></td><td colspan="2">Average 0.63</td><td></td><td></td></tr>
<tr><td colspan="6">Main idea prompt</td></tr>
<tr><td>Blaha, 1979</td><td>0.70*[d]</td><td>0.88*</td><td></td><td></td><td></td></tr>
<tr><td>Dreher & Gambrell, 1985</td><td></td><td></td><td>0.00 (est)</td><td></td><td></td></tr>
<tr><td>Lonberger, 1988 (RT)</td><td></td><td></td><td>1.24*</td><td></td><td></td></tr>
<tr><td>Ritchie, 1985</td><td></td><td></td><td>0.00 (est)</td><td></td><td></td></tr>
<tr><td>Wong & Jones, 1982</td><td></td><td></td><td>Reg. 0.00 (est)
LD 0.50* (est)
Average 0.25</td><td></td><td></td></tr>
<tr><td colspan="6">Question type prompt</td></tr>
<tr><td>Dermody, 1988 (RT)</td><td>–0.32[f]</td><td></td><td></td><td></td><td>3.37*</td></tr>
<tr><td>Labercane & Battle, 1987 (RT)</td><td>0.36[a] (est)</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Smith, 1977</td><td>0.00[g] (est)</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Story grammar prompt</td></tr>
<tr><td>Nolte & Singer, 1985</td><td></td><td>1.01*</td><td></td><td></td><td></td></tr>
<tr><td>Short & Ryan, 1984</td><td></td><td></td><td>1.22* (est)</td><td>1.05* (est)</td><td></td></tr>
<tr><td colspan="6">No prompt</td></tr>
<tr><td>Helfeldt & Lalik, 1976</td><td>0.84*[h]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Manzo, 1969</td><td>0.14[a]</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Simpson, 1989</td><td>–0.25[i]</td><td></td><td></td><td></td><td></td></tr>
</table>

Note. Medians used when scores combined. A study marked "(RT)" used reciprocal teaching as an instructional approach.

[a]Gates-MacGinitie Reading Tests. [b]Developmental Readiness Test. [c]Metropolitan. [d]Nelson-Denny Reading. [e]Illinois State Reading Assessment. [f]Stanford Diagnostic Reading Test. [g]Iowa Tests of Basic Skills. [h]Van-Wagenen Analytic Reading Scales. [i]Reading Comprehension subtest of the California Achievement Test.

*Significant.

APPENDIX C

Three indicators of quality

Study	Provided modeling and guided practice	Used comprehension probes	Assessed student learning of strategy
	Signal words prompt		
Brady, 1990 (RT)	Yes	Yes	
Cohen, 1983	Yes		Yes*
Davey & McBride, 1986	Yes		Yes*
Lysynchuk et al., 1990 (RT)	Yes	Yes	Yes (ns)
MacGregor, 1988	Yes		Yes (but no comp. group; not test)
Palincsar, 1987 (RT)	Yes	Yes	Yes (ns)
Palincsar & Brown, 1984 (RT)	Yes	Yes	Yes (ns)
Taylor & Frye, 1988 (RT)	Yes		Yes (ns)
Williamson, 1989 (RT)	Yes		
	Generic questions or question stems prompt		
King, 1989	Yes	Yes	
King, 1990	Yes		
King, 1992	Yes		No test; stated students reached proficiency in training; no data given
Weiner, 1977	Yes		
	Main idea prompt		
Blaha, 1979	Yes		
Dreher & Gambrell, 1985	Yes		Yes; no stat. analysis
Lonberger, 1988 (RT)	Yes		Yes (ns)
Ritchie, 1985	Yes		Yes*
Wong & Jones, 1982	Yes		Yes no compar. with control
	Question types prompt		
Dermody, 1988 (RT)	Yes	Yes	Yes*
Labercane & Battle, 1987 (RT)	Yes		
Smith, 1977	Yes		Yes*
	Story grammar prompt		
Nolte & Singer, 1985	Yes		
Short & Ryan, 1984	Yes		
	No prompt		
Helfeldt & Lalik, 1976	Yes		
Manzo, 1969	Yes		Yes* on higher-level ques.
Simpson, 1989	Yes		

*Results were statistically significant when control group was compared to treatment group on the assessment of strategy use.

APPENDIX D

Overall effect size table

Study	Effect size
Reciprocal teaching	
Standardized	
Brady, 1990	0.36
Lysynchuk et al., 1990	0.55
Taylor & Frye, 1988	0.07
Williamson, 1989	0.32
Dermody, 1988	–0.32
Labercane & Battle, 1987	0.36
Median effect size	**0.34**
Experimental	
Brady, 1990	0.87*
Lysynchuk et al., 1990	0.68*
Palincsar, 1987	0.88*
Palincsar & Brown, 1984	1.00*
Taylor & Frye, 1988	0.85
Dermody, 1988	3.37*
Lonberger, 1988	1.24*
Median effect size	**0.88**
Other treatments	
Standardized	
Cohen, 1983	0.57*
MacGregor, 1988	0.35
Smith, 1977	0.00
Blaha, 1979	0.70*
Simpson, 1989	–0.25
Helfeldt & Lalik, 1976	0.84*
Manzo, 1969	0.14
Median effect size	**0.35**
Experimental	
Cohen, 1983	0.57*
Davey & McBride, 1986	0.77*
Dreher & Gambrell, 1985	0.00
Ritchie, 1985	0.00
Blaha, 1979	0.88*
Wong & Jones, 1982	0.25
King, 1989	1.37*
King, 1990	1.70*
King, 1992	0.87*
Weiner, 1977	0.63*
Short & Ryan, 1984	1.14*
Nolte & Singer, 1985	1.01*
Median effect size	**0.82**

Note. Overall median effect size for all self-questioning studies = 0.61.

Estimated effect size could be determined through *p*-value when actual *t* or *F* was not available. For a study for which we could not calculate an effect size or an estimated effect size because of 3 or more treatments, we assigned an effect size of 0; this provided a more conservative median effect size for nonsignificant studies than when they were assigned the median effect size of all nonsignificant studies (0.30). We assigned Wong & Jones (1982) an effect size of 0.25, which was the average of the assigned 0 for the nonsignificant result and a 0.5 for an estimated effect size calculated for the significant result.

*Significant.

References

Adams, A., Carnine, D., & Gersten, R. (1982). Instructional strategies for studying content area texts in the intermediate grades. *Reading Research Quarterly, 18,* 27–55.

Alvermann, D. E. (1981). The compensatory effect of graphic organizers on descriptive text. *Journal of Educational Research, 75,* 44–48.

Anderson, V., & Roit, M. (1993). Planning and implementing collaborative strategy instruction with delayed readers in Grades 6–10. *Elementary School Journal, 93,* 134–148.

Andre, M. D. A., & Anderson, T. H. (1979). The development and evaluation of a self-questioning study technique. *Reading Research Quarterly, 14,* 605–623.

Armbruster, B., Anderson, T., & Ostertag, B. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly, 22,* 332–345.

Armbruster, B. B. (1984). The problem of "inconsiderate" text. In G. G. Duffy, L. R. Roehler, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (pp. 202–217). New York: Longman.

Bandura, A. (1969). *Principles of behavior modification.* New York: Holt, Rinehart, and Winston.

Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly, 20,* 93–115.

Billingsley, B. S., & Wildman, T. M. (1984). Question generation and reading comprehension. *Learning Disability Research, 4,* 36–44.

Blaha, B. A. (1979). *The effects of answering self-generated questions on reading.* Unpublished doctoral dissertation, Boston University School of Education.

Brady, P. L. (1990). *Improving the reading comprehension of middle school students through reciprocal teaching and semantic mapping strategies.* Unpublished doctoral dissertation, University of Alaska.

Braun, C., Rennie, A., & Labercane, G. (1985, December). *A conference approach to the development of metacognitive strategies.* Paper presented at the annual meeting of the National Reading Conference, San Diego, CA. (ERIC Document Reproduction Service No. ED 270 734)

Cohen, R. (1983). Students generate questions as an aid to reading comprehension. *Reading Teacher, 36,* 770–775.

Collins, A., Brown, J. S., & Newman, S. E. (1990). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Craig, F., & Lockhart, R. S. (1972). Levels of processing: Framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Davey, B., & McBride, S. (1986). Effects of question-generation on reading comprehension. *Journal of Educational Psychology, 78,* 256–262.

Dermody, M. (1988, February). *Metacognitive strategies for development of reading comprehension for younger children.* Paper presented at the annual meeting of the American Association of Colleges for Teacher Education. New Orleans, LA.

Dreher, M. J., & Gambrell, L. B. (1985). Teaching children to use a self-questioning strategy for studying expository text. *Reading Improvement, 22,* 2–7.

Duffy, G., Roehler, L., Sivan, E., Rackliffe, G., Book, C., Meloth, M., Varus, L., Wasselman, R., Putman, J., & Bassiri, D. (1987). The effects of explaining reading associated with using reading strategies. *Reading Research Quarterly, 22,* 347–367.

Durkin, D. (1979). What classroom observations reveal about reading comprehension. *Reading Research Quarterly, 14,* 518–544.

Englert, C. S., & Raphael, T. E. (1989). Developing successful writers through cognitive strategy instruction. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 1). Newark, NJ: JAI Press.

Frase, L. T., & Schwartz, B. J. (1975). Effect of question production and answering on prose recall. *Journal of Educational Psychology, 67,* 628–635.

Gagne, R. M. (1977). *The conditions of learning* (3rd ed.). New York: CBS College Publishing.

Garcia, G. E., & Pearson, P. D. (1990). *Modifying reading instruction to maximize its effectiveness for all students* (Tech. Rep. No. 489). Champaign, IL: University of Illinois, Center for the Study of Reading.

Gilroy, A., & Moore, D. (1988). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities with ten primary school girls. *Educational Psychology, 8,* 41–49.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Good, T. L., & Brophy, J. E. (l986). School effects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 570–602). New York: Macmillan.

Hare, V. C., & Borchart, K. M. (1985). Direct instruction of summarization skills. *Reading Research Quarterly, 20,* 62–78.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1991) *A practical guide to modern methods of meta-analysis.* Washington, DC: National Association of Research in Science Teaching.

Helfeldt, J. P., & Lalik, R. (1976). Reciprocal student-teacher questioning. *Reading Teacher, 33,* 283–287.

Heller, J. I., & Hungate, H. N. (1985). Implications for mathematics instruction of research on scientific problem solving. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives.* Hillsdale, NJ: Erlbaum.

Heller, J. I., & Reif, F. (1984). Prescribing effective human problem-solving processes: Problem description in physics. *Cognition and Instruction, 1,* 177–216.

Hillocks, G., Jr. (1987). Synthesis of research on teaching writing. *Educational Leadership,* 44, 71–82.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology, 14,* 366–381.

King, A. (1990). Improving lecture comprehension: Effects of a metacognitive strategy. *Applied Educational Psychology, 5,* 331–346.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29,* 303–325.

King, A. (1994). Autonomy and question asking: The role of personal control in guided student-generated questioning. *Learning and Individual Differences, 6,* 163–185.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative-questioning on children's knowledge construction. *Journal of Experimental Education, 6,* 127–148.

Knudson, R. E. (1988). The effects of writing highly structured versus less structured lessons on student writing. *Journal of Educational Research, 81,* 365–368.

Labercane, G., & Battle, J. (1987). Cognitive processing strategies, self-esteem, and reading comprehension of learning disabled students. *Journal of Special Education, 11,* 167–185.

Larkin, J. H., & Reif, F. (1976). Analysis and teaching of a general skill for studying scientific text. *Journal of Educational Psychology, 72,* 348–350.

Lonberger, R. B. (1988). *The effects of training in a self-generated learning strategy on the prose processing abilities of fourth and sixth graders.* Unpublished doctoral dissertation, State University of New York at Buffalo.

Lysynchuk, L., Pressley, M., & Vye, G. (1990). Reciprocal instruction improves reading comprehension performance in poor grade school comprehenders. *Elementary School Journal, 40,* 471–484.

MacGregor, S. K. (1988). Use of self-questioning with a computer-mediated text system and measures of reading performance. *Journal of Reading Behavior, 20,* 131–148.

Manzo, A. V. (1969). *Improving reading comprehension through reciprocal teaching.* Unpublished doctoral dissertation, Syracuse University.

Meichenbaum, D. (1977). *Cognitive behavior modification: An integrative approach.* New York: Plenum.

Nolte, R. Y., & Singer, H. (1985). Active comprehension: Teaching a process of reading comprehension and its effects on reading achievement. *The Reading Teacher, 39,* 24–31.

Palincsar, A. S. (1987, April). *Collaborating for collaborative learning of text comprehension.* Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 2,* 117–175.

Paris, S. C., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76,* 1239–1252.

Pressley, M. J., Burkell, J., Cariglia-Bull, T., Lysynchuk, L., McGoldrick, J. A., Schneider, B., Snyder, B. L., Symons, S., Woloshyn, V. E. (1990). *Cognitive strategy instruction that really improves children's academic performance.* Cambridge, MA: Brookline Books.

Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J. A., & Kurita, J. A. (1989). Strategies that improve children's memory and comprehension of text. *Elementary School Journal, 90,* 3–32.

Raphael, T. E., & Pearson, P. D. (1985). Increasing student awareness of sources of information for answering questions. *American Educational Research Journal, 22,* 217–237.

Raphael, T. E., & Wonnacott, C. A. (1985). Meta-cognitive training in question-answering strategies: Implementation in a fourth grade developmental reading program. *Reading Research Quarterly, 20,* 282–296.

Rickards, J. P., & Denner, P. P. (1979). Depressive effects of underlining and adjunct questions on children's recall of text. *Instructional Science, 8,* 80–91.

Ritchie, P. (1985). The effects of instruction in main idea and question generation. *Reading Canada Lecture, 3,* 139–146.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmillan.

Scardamalia, M., & Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2. Research and open questions* (pp. 563–577). Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* New York: Academic Press.

Short, E. J., & Ryan, E. B. (1984). Metacognitive differences between skilled and less-skilled readers: Remediating deficits through story grammar and attribution training. *Journal of Educational Psychology, 76,* 225–235.

Simpson, P. S. (1989). *The effects of direct training in active comprehension on reading achievement, self-concepts, and reading attitudes of at-risk sixth grade students.* Unpublished doctoral dissertation, Texas Tech University.

Singer, H. (1978). Active comprehension: From answering to asking questions. *Reading*

Teacher, 31, 901–908.
Singer, H., & Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation of complex short stories. *Reading Research Quarterly, 17,* 166–186.
Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research, 57,* 175–215.
Smith, N. J. (1977). *The effects of training teachers to teach students at different reading ability levels to formulate three types of questions on reading comprehension and question generation ability.* Unpublished doctoral dissertation, University of Georgia.
Taylor, B. M., & Frye, B. J. (1992). Comprehension strategy instruction in the intermediate grades. *Reading Research and Instruction, 92,* 39–48.
Weiner, C. J. (1978, March). *The effect of training in questioning and student question-generation on reading achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Ontario, Canada. (ERIC Document Reproduction Service No. ED 158 223)
Weinstein, C. E. (1978). Teaching cognitive elaboration learning strategies. In H. F. O'Neal (Ed.), *Learning strategies.* New York: Academic Press.
Williamson, R. A. (1989). *The effect of reciprocal teaching on student performance gains in third grade basal reading instruction.* Unpublished doctoral dissertation, Texas A&M University.
Wong, B. Y. L. (1985). Self-questioning instructional research: A review. *Review of Educational Research, 55,* 227–268.
Wong, B. Y. L., & Jones, W. (1982). Increasing metacomprehension in learning disabled and normally achieving students through self-questioning training. *Learning Disability Quarterly, 5,* 228–239.
Wong, B. Y. L., Wong, W., Perry, N., & Sawatsky, D. (1986). The efficacy of a self-questioning summarization strategy for use by underachievers and learning-disabled adolescents. *Learning Disability Focus, 2,* 20–35.
Wood, D. J., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17,* 89–100.

Authors

BARAK ROSENSHINE is Professor, Department of Educational Psychology, College of Education, University of Illinois at Urbana, 1310 South Sixth, Champaign, IL 61820. He specializes in classroom instruction and cognitive strategy research.
CARLA MEISTER is a doctoral candidate, Department of Educational Psychology, College of Education, University of Illinois at Urbana, 1310 South Sixth, Champaign, IL 61820; carlameist@aol.com. She specializes in learning and instruction, and cognitive strategies in reading and writing.
SAUL CHAPMAN is a doctoral candidate, Department of Educational Psychology, College of Education, University of Illinois at Urbana, 1310 South Sixth, Champaign, IL 61820. He specializes in counseling psychology.

Received April 11, 1995
Revision received August 4, 1995
Accepted August 20, 1995
Withdrawn October 17, 1995
Resubmitted February 7, 1996
Accepted March 27, 1996