

Review of Educational Research

<http://rer.aera.net>

Reciprocal Teaching: A Review of the Research

Barak Rosenshine and Carla Meister

REVIEW OF EDUCATIONAL RESEARCH 1994 64: 479

DOI: 10.3102/00346543064004479

The online version of this article can be found at:

<http://rer.sagepub.com/content/64/4/479>

Published on behalf of



American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *Review of Educational Research* can be found at:

Email Alerts: <http://rer.aera.net/alerts>

Subscriptions: <http://rer.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

Citations: <http://rer.sagepub.com/content/64/4/479.refs.html>

>> [Version of Record](#) - Jan 1, 1994

[What is This?](#)

Reciprocal Teaching: A Review of the Research

Barak Rosenshine

Carla Meister

University of Illinois at Urbana

Reciprocal teaching is an instructional procedure designed to teach students cognitive strategies that might lead to improved reading comprehension. The learning of cognitive strategies such as summarization, question generation, clarification, and prediction is supported through dialogue between teacher and students as they attempt to gain meaning from text. This article is a review of sixteen studies on reciprocal teaching, which include published studies found in journal articles and unpublished studies indexed in Dissertation Abstracts International. All the studies included in this review were quantitative in methodology. When standardized tests were used to assess comprehension, the median effect size, favoring reciprocal teaching, was .32. When experimenter-developed comprehension tests were used, the median effect size was .88. We also discuss the role of cognitive strategies in enhancing comprehension, the strategies that were most helpful, instructional approaches for teaching cognitive strategies, the quality of the dialogue during reciprocal teaching, and suggestions for future research and practice.

Reciprocal teaching is an instructional approach that features “guided practice in applying simple, concrete strategies to the task of text comprehension” (Brown & Palincsar, 1989, p. 413). Reciprocal teaching was first described by Palincsar (1982) and Palincsar and Brown (1984), and the description was extended in their later articles, particularly in Palincsar (1986) and Brown and Palincsar (1989). This is a review of the studies that have used reciprocal teaching as an instructional procedure to improve student comprehension of text.

Until the late 1970s, students were seldom taught cognitive strategies that could assist them in reading. 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 a great deal of instructional time

This research was sponsored by the Bureau of Educational Research, College of Education, University of Illinois at Urbana–Champaign. Special thanks are extended to those who assisted us during the writing of this review: Patricia Alexander, Philip Brady, Russell Burnett, Courtney Cazden, Saul Chapman, Lyn Corno, Margaret Dermody, Mark Gall, John Helfeldt, Fran Lehr, Linda Lysynchuk, Anthony Manzo, Carol Moore, Annemarie Palincsar, Mike Pressley, Rebecca Rich, Bill Rupley, Jo Shepard, Bob Slavin, Donna Weisman, Philip Zodhiates, and two of the *RER* reviewers.

asking students questions, but they spent little time teaching students comprehension strategies they could use to answer the questions. Duffy, Lanier, and Roehler (1980) noted a similar lack of comprehension instruction in elementary classrooms:

There is little evidence of instruction of any kind. Teachers spend most of their time assigning activities, monitoring to be sure the pupils are on task, directing recitation sessions to assess how well children are doing and providing corrective feedback in response to pupil errors. Seldom does one observe teaching in which a teacher presents a skill, a strategy, or a process to pupils, shows them how to do it, provides assistance as they initiate attempts to perform the task and assures that they can be successful. (p. 4)

In the late 1970s, investigators began to teach students specific cognitive strategies, such as question generation or summarization, that they could use to help improve their reading comprehension (Paris, Cross, & Lipson, 1984; Raphael & Pearson, 1985; Alvermann, 1981). Cognitive strategy instruction has also been taught in mathematics problem solving (Schoenfeld, 1985), physics problem solving (Larkin & Reif, 1976), and writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985). Reciprocal teaching is in this tradition of cognitive strategy instruction.

Reciprocal teaching refers to a set of learning conditions in which children “first experience a particular set of cognitive activities in the presence of experts, and only gradually come to perform these functions by themselves” (Brown & Palincsar, 1989, p. 123). In reciprocal teaching (a) the focus is upon teaching students specific, concrete, comprehension-fostering strategies which they can apply to the reading of new text, and (b) this instruction takes place primarily in the context of a dialogue between the teacher and the students.

In reciprocal teaching, as developed by Palincsar and Brown (1984), students read a passage of expository material, paragraph by paragraph. During the reading they learn and practice four reading comprehension strategies: generating questions, summarizing, attempting to clarify word meanings or confusing text, and predicting what might appear in the next paragraph. During the early stages of reciprocal teaching, the teacher assumes the major responsibility for instruction by explicitly modeling the process of using these strategies on a selection of text. After the teacher has modeled, the students practice the strategies on the next section of text, and the teacher supports each student’s participation through specific feedback, additional modeling, coaching, hints, and explanation. The teacher adjusts the difficulty of the task according to the current level of the student, as described by Palincsar and Brown (1984):

The teacher models and explains, relinquishing part of the task to novices only at the level each one is capable of negotiating at any one time. Increasingly, as the novice becomes more competent, the teacher increases her demands, requiring participation at a slightly more challenging level. (p. 13)

During this guided practice the teacher invites students to initiate discussion and to react to other students’ statements. Students’ participation can include (a) elaborating or commenting on another student’s summary, (b) suggesting other questions, (c) commenting on another’s predictions, (d) requesting clarification of material they did not understand, and (e) helping to resolve misunderstandings.

The teacher supports the students by rephrasing or elaborating on their answers, statements, and questions. In the course of this guided practice, there is a gradual shift in responsibility from the teacher doing much of the work to the child taking over the major thinking role while the teacher observes and helps only when needed.

At this point, the practice becomes a dialogue: one student asks questions, another answers, and a third comments on the answer; one student summarizes and another comments on or helps to improve the summary; one student identifies a difficult word and the other students help to infer the meaning and give reasons for the inferences they made. The emphasis throughout is on cooperative effort by teacher and students to bring meaning to the ideas in the text, rather than merely restating the words. In addition, during the dialogue, students are provided instruction in why, when, and where such activities should be applied to new text.

Thus, reciprocal teaching has two major features. The first is instruction and practice of four *comprehension-fostering* strategies: question generation, summarization, prediction, and clarification. The second consists of the use of the reciprocal teaching dialogue as a vehicle for learning and practicing these four strategies.

The process of gradual introduction of a skill by a teacher who provides assistance to students as they practice is similar to the guided practice described by Hunter (1982), Good and Grouws (1979), and Rosenshine and Stevens (1986). In reciprocal teaching, however, much greater emphasis is placed on encouraging students to provide instructional support for each other.

Selecting studies. We located the intervention studies in reciprocal teaching by searching the ERIC and Dissertation Abstracts International databases, as well as programs for the annual meetings of the American Educational Research Association. We included only studies that (a) explicitly used the words *reciprocal teaching*, (b) referenced the work of Palincsar and Brown (1984), and (c) contained both experimental and control groups in which the authors randomly assigned students to the two groups or determined the two groups to be similar on initial measures of reading comprehension.

We did not include the earlier studies by Manzo (1969) and Helfeldt and Lalik (1976), which were in the Reciprocal Questioning tradition, because these studies taught only a single strategy, question generation, and did not provide the support, the scaffolded instruction, and the dialogue that occurs in reciprocal teaching. We also did not include the study by Nolte and Singer (1985), even though they provided explicit modeling, feedback, and student practice, because they taught only a single strategy, question generation, and did not claim to be in the reciprocal teaching tradition. Those studies that taught the strategy of question generation but are outside the reciprocal teaching approach are discussed in the Discussion section.

We located 16 studies that meet the above three criteria. In 12 of the studies, students were taught the four comprehension-fostering strategies introduced by Palincsar and Brown (1984); in the other four studies, 2, 3, or 10 cognitive strategies were taught. We decided to include all 16 studies in our analysis, but also to conduct separate analyses for studies that used Palincsar and Brown's four strategies and for studies that used a different number of strategies.

Four of the 16 studies were published (Labercane & Battle, 1987; Lysynchuk, Pressley, & Vye, 1990; Palincsar & Brown, 1984; Taylor & Frye, 1992). The remaining 12 were dissertations and/or papers presented at meetings of the American Educational Research Association or the National Reading Conference. In all cases involving dissertations, we obtained and studied the complete dissertations. All studies used expository material during the reciprocal teaching dialogues.

Two forms of reciprocal teaching. Although the reciprocal teaching dialogue is common to all the studies, two forms of reciprocal teaching evolved in the course of the work of Palincsar and Brown: *reciprocal teaching only* (RTO) and *explicit teaching before reciprocal teaching* (ET-RT). The two forms differ in how and when the initial instruction in the cognitive strategies takes place.

In the first form, exemplified by Palincsar and Brown's original 1984 study, all modeling and instruction on how to develop and apply the four cognitive strategies takes place *during* the dialogues. It is during the dialogues that the students are given procedural prompts, words such as *how* or *when*, to help them generate questions. It is during the dialogues that the students are given procedures to help them develop a summary or clarify the meaning of a word.

We refer to this form of reciprocal teaching, described in the original Palincsar and Brown (1984) study, as reciprocal teaching only (RTO). We believe that 7 of the 16 studies in this review fit into the RTO category: Fischer Galbert (1989), Jones (1987), Labercane and Battle (1987), Padron (1985), Palincsar and Brown (1984), Rich (1989), and Rush and Milburn (1988). Details on all 16 studies are presented in Appendices A and B.

The following examples, taken from the Palincsar and Brown (1984) study, illustrate how the teacher provides prompts, models, cues, and feedback on the use of cognitive strategies during the dialogues.

(Questioning)

T: What would be a good question about pit vipers that starts with the word "why?"

S: (No response)

T: How about, "Why are the snakes called pit vipers?"

T: That's good. Keep going.

S: How do spinner's mate is smaller than. How am I going to say that?

T: Take your time with it. You want to ask a question about the spinner's mate and what he does, beginning with the word "how."

S: How do they spend most of his time sitting?

T: You're very close. The question would be "How does spinner's mate spend most of his time?" Now you ask it.

S: How does spinner's mate spend most of his time?

S: Snakes' backbones can have as many as 300 vertebrae almost __ times as many as humans.

T: Not a bad beginning, but I would consider that a question about a detail. See if the next time you can find a main idea question and begin your question with a question word—how, why, when. . . .

(Summarizing)

T: That was a fine job, Ken, but I think there might be something to add to our summary. There is more information that I think we need to include. This paragraph is mostly about what?

S: The third method of artificial evaporation.

As the dialogues continue, students in the group progressively take on more responsibility for carrying out the dialogue. Students in the group begin to provide models, hints, and prompts to each other, as well as feedback regarding the use of the strategies. Although the teacher is not absent from these later discussions, the role of the teacher shifts to that of a sympathetic coach. (pp. 138, 142)

A second form of reciprocal teaching was found in three subsequent studies by Palincsar and her associates (Palincsar, 1987; Palincsar, Brown, & Martin, 1987; Palincsar, David, Winn, Stevens, & Brown, 1990). In these studies, the students were first introduced to the four strategies during three to six traditional lessons that were conducted *before* the dialogues began. These lessons were developed to “introduce the students to the ‘language’ of reciprocal teaching by providing direct instruction in each strategy” (Palincsar et al., 1987, pp. 235, 238; Palincsar et al., 1990, p. 15; Brown & Palincsar, 1989, p. 33). In these lessons, “students were introduced to the strategies, one by one, using worksheet activities led by the classroom teachers and conducted on a whole class basis” (Palincsar et al., 1987, p. 238). Separate lessons were devoted to instruction in question generation, summarization, clarification, and prediction. An example of one of these introductory lessons is reproduced in Appendix C.

The reciprocal teaching dialogues and the guided practice that occurred during the dialogues were found in both forms of reciprocal teaching, but this second form differs from the first in that it provides explicit instruction on the four strategies *before* beginning the dialogues. We refer to this second form of reciprocal teaching as explicit teaching before reciprocal teaching (ET-RT). We placed nine studies in this category: Brady (1990), Dermody (1988), Levin (1989), Lonberger (1988), Lysynchuk et al. (1990), Palincsar (1987), Shortland-Jones (1986), Taylor and Frye (1992), and Williamson (1989).

Thus, these two related forms of reciprocal teaching appeared in the studies by Palincsar, Brown, and their associates, and also in studies conducted by the authors of the other studies in this review. Although Palincsar and Brown did not present any theoretical explanation for this change from RTO to ET-RT, these two variations seemed sufficiently different to suggest exploring whether they yielded different results. Thus, we conducted some separate analyses of the results for the RTO and the ET-RT studies.

Theoretical Rationale

Brown and Palincsar (1989) have noted that reciprocal teaching is explained by three related theories of guided learning: Vygotsky’s zone of proximal development (Vygotsky, 1978), proleptic teaching (Wertsch & Stone, 1979; Rogoff & Gardner, 1984), and expert scaffolding (Wood, Bruner, & Ross, 1976).

Vygotsky (1978) observed that a child has two developmental levels. One is the *actual developmental level*, the level at which children can independently deal with tasks. The other is the *level of potential development*, or the level at which a child can solve a problem with the assistance of a teacher or in collabora-

tion with other children. The *zone of proximal development* is the area between the actual developmental level of the child and the level of potential development (Vygotsky, 1978, pp. 85–86), and is the area within which instruction can take place.

Vygotsky believed that one does not have to wait until a child is developmentally ready before beginning instruction. Rather, he emphasized the role of instruction in *fostering* development. Vygotsky wrote, “In the child’s development, imitation and instruction play a major role. . . . Therefore, the only good kind of instruction is that which marches ahead of development and leads it; it must be aimed not so much at the ripe as at the ripening functions” (Vygotsky, 1962, p. 104).

Unfortunately, Vygotsky did not provide much information on how one might instruct children in their zone of proximal development (see Palincsar et al., 1989, p. 2). For one source of ideas about instruction, Palincsar and Brown (1984) and Brown and Palincsar (1989) turned to the procedures found in apprenticeship instruction in areas such as weaving (Greenfield, 1984) and tailoring (Lave, 1977). This type of instruction has been termed *proleptic teaching* (Wertsch & Stone, 1979), or teaching “in anticipation of competence.”

In these settings, it is the adult who is responsible for getting the job done, with the child participating first as a spectator, then as a novice responsible for very little of the actual work. As the apprentices become more experienced and capable of performing more complex aspects of the task, aspects that have been modeled by the adults time and time again, they are ceded greater and greater responsibility until they become experts themselves. Within these systems of tutelage, novices learn about the task at their own rate, in the presence of experts, participating only at a level they are capable of fulfilling at any point in time. (Brown & Palincsar, 1989, p. 410)

The experts begin by modeling the task for the child. Then they assist the child and gradually cede more and more responsibility to the child until the child can do the task independently (Brown & Palincsar, 1989). Palincsar and Brown have written that this instructional pattern appears in the work of Rogoff and Gardner (1984), Vygotsky (Palincsar & Brown, 1984, p. 123) and Bruner (Brown & Palincsar, 1985, p. 13).

Brown and Palincsar (1989) also present a system of instruction they call *expert scaffolding*. In expert scaffolding, the expert acts as a guide, shaping the learning efforts of the novices and providing support for the learning until it is no longer needed. “The metaphor of a scaffold captures the idea of an adjustable and temporary support that can be removed when no longer necessary” (p. 411). The first use of the term “scaffolding” was by Wood et al. (1976) to refer to the instructional process whereby an adult controls “those elements of the task that are initially beyond the learner’s capacity, thus permitting the learner to concentrate upon and complete only those elements that are within his range of competence” (p. 90). Several scaffolding procedures identified by Wood et al. include reducing the complexity of the task to manageable limits, maintaining student interest, marking critical features, and demonstrating solutions when the learner can recognize them. Although Vygotsky never used the term “scaffolds,” he conceptualized the technique when he wrote the following:

We might study the development of memorizing in children by making available to them new means for solving the given task and then observing the degree and character of their problem solving efforts. (1978, p. 74)

In sum, Palincsar and Brown (1984) rooted reciprocal teaching instruction in the theoretical work of Vygotsky (1978), Rogoff and Gardner (1984), and Wood et al. (1976). Vygotsky provided the concepts of level of potential development, zone of proximal development, and children providing support for one another. Both Rogoff and Gardner (1984) and Wood et al. (1976) contributed ideas for expert modeling, expert support as the learner begins the task, and gradual fading of this support.

It is possible that Palincsar and Brown developed their reciprocal teaching procedures without this theoretical foundation, and after the study was completed found that Vygotskyian theory could be used to describe their successful practice. We suggest this possibility because there was no reference to Vygotsky, Wood, Bruner, Rogoff and Gardner, or Wertsch and Stone in the review-of-literature section of Palincsar's (1982) dissertation. (The dissertation provided the data for the Palincsar and Brown, 1984 article). In her dissertation, Palincsar referred to studies on cognitive strategy instruction (e.g., Bird, 1980; Brown, Campione, & Day, 1981; Dansereau, 1980) as well as research on teacher effects (Anderson, Evertson, & Brophy, 1979; Rosenshine & Berliner, 1978). However, she did not cite Vygotsky, Rogoff, Wood, or Bruner (see Palincsar, 1982, pp. 34–35).

In 1983, a report on Palincsar's dissertation study was issued from the Center for the Study of Reading (Palincsar & Brown, 1983). Again, in that report, there were no references to the theoretical positions of Vygotsky, Rogoff, Wood, or Bruner. Rather, the major references were to experimental studies in which cognitive strategies were explicitly taught to learners, particularly the work of Andre and Anderson (1978–79); Brown, Campione, and Day (1981); Dansereau et al. (1979); Meichenbaum (1977); and Wong and Jones (1982).

In educational research, theory and practice interact so that theory supports practice and practice supports theory. In this case, the practice, reciprocal teaching, may have been developed first and the investigators may have found later that Vygotskyian theory, scaffolds, and proleptic teaching supported the practice and helped to explain its success. The result of this interaction can lead to a refinement of both practice and theory. Whatever the source, the concept of reciprocal teaching, as developed by Palincsar and Brown, provides us with both a theoretical rationale and a well written exposition of the concept of guided practice. This guided practice includes the instructional concepts of expert modeling, expert support as the learner begins the task, children providing support for one another, and gradual fading of support as the child develops competence.

Quality of the Studies

We attempted to evaluate the quality of the studies by looking at design of the study, assessments of student learning, and assessments of the quality of the reciprocal teaching dialogues. We then used the results of these analyses to classify each study as high, medium, or low in quality. All criteria used to determine the quality of each study, the information on each study, and the overall quality ratings are presented in Table 1. All evaluations were made after reading the complete study. We obtained and studied copies of all dissertations.

TABLE 1
Criteria for determining quality of study

Study	Person providing instruction	Comp. control group	Transfer test	Comprehension probes during study	Assessed student learning of strategy	Teacher implem. observed	Adequate instruction based on observation	Adequate instruction based on description	Quality rating
Brady, 1990	Investigator	Yes	Yes	Yes	Yes (audiotaped use of all four strategies)			Yes	Middle
Dermody, 1988	Investigator	Yes	Yes	Yes				Yes	High
Fischer Galbert, 1989	Teachers	Yes	Yes					Yes	Middle
Jones, 1987	Investigator	Yes	Yes					Yes	Middle
Labercane & Battle, 1987	Investigator	Yes	Yes					Yes	Middle
Levin, 1989	Teachers	Yes	Yes			Yes		Yes	Middle
Lonberger, 1988	No indication (assumed investigator)	Yes	Yes				Yes	Yes	Middle
Lysynchuk et al., 1990	Investigator	Yes	Yes	Yes	Yes			Yes	High
Padron, 1985	Investigator	Yes	Yes					Uncertain	Uncertain
Palincsar, 1987	Teachers	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Palincsar & Brown, 1984	Investigator	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High
Rich, 1989	Investigator	Yes	Yes					Yes	Middle
Rush & Milburn, 1988	Trained students	Yes	Yes					Uncertain	Uncertain
Shortland-Jones, 1986	Teachers	Yes	Yes		Yes	Yes	Yes	Yes	High
Taylor & Frye, 1992	Teachers	Yes	Yes		Yes	Yes	Yes	Yes	High
Williamson, 1989	Teachers	Yes	Yes			Yes	No	No	Low

Design of the Study

We include in this review only studies that have comparable experimental and control groups. We exclude studies without comparable control groups, and studies that lack control groups altogether. For example, we eliminated studies in which only poor readers were in the experimental group and average readers were in the control group (e.g., Gilroy & Moore, 1988). We excluded Palincsar and Brown's (1984) Study 2, even though it used a time-series design, because it lacked a control group. However, we accepted studies that had a comparable control group even when those students were not randomly selected. Thus, we accepted Palincsar and Brown's (1984) Study 1 because the pretest scores of their four groups were comparable, even though the control groups were selected from similar students some months after the treatment was completed (see Palincsar & Brown, 1983, p. 25). All studies that were included assessed student comprehension by asking students to read new passages that had not been used during the training.

Assessments of Student Learning

We also checked to see (a) whether the studies assessed the quality of student comprehension during the study and (b) whether they assessed student learning of the strategies. Palincsar and Brown (1984) and Palincsar (1987) assessed student comprehension periodically, during the study, by giving the students new passages and asking them to answer ten short-answer questions. The results of these quizzes were then plotted as a way of monitoring student progress. Palincsar developed these passages and tests. Surprisingly, this sensible procedure was used in only 3 of the remaining 14 studies. Brady (1990) used the Palincsar materials in his study; Lysynchuk et al. (1990) and Dermody (1988) developed their own comprehension assessment materials (see Table 1).

Palincsar and Brown (1984) and Palincsar (1987) also assessed student learning of the strategy of question generation by giving students new passages and asking them to generate "questions a teacher might ask." Five other studies also tested students on their ability to generate questions. These studies are listed in Table 1 in the column "Assessed student learning of strategy."

Assessments of the Quality of the Reciprocal Teaching Dialogues

The most critical part of reciprocal teaching is the quality of its dialogue (Palincsar, 1986). Unfortunately, most authors did not provide much information that we could use to evaluate the quality of the dialogue in each study. Nonetheless, we looked at the investigators' results for (a) direct observation of the quality of instruction, and (b) evaluation of the quality of the student questions and summaries during the dialogues. We then attempted to assess each investigator's knowledge of reciprocal teaching.

Direct observation of instruction. We searched the studies to see whether the quality of the reciprocal teaching was assessed; our findings are presented in Table 1. It was difficult to develop criteria to evaluate the quality of the dialogue because Palincsar and Brown did not provide a list of specific instructional procedures that represent quality implementation of reciprocal teaching. Although they wrote that "the first author checked weekly on the teacher-directed sessions

to see if the intervention was being conducted properly” (1984, p. 158), they gave no information about how these determinations were made. When Palincsar conducted another study in 1987, she wrote that “the raters noted that each of the teachers indicated fidelity to the reciprocal teaching procedure” (p. 18), but the specific criteria used to judge that quality were not given. As a result, subsequent researchers had to develop their own criteria.

In seven studies, the investigators trained classroom teachers to provide reciprocal teaching instruction to students, and the investigators assessed the level of teacher performance in five of these seven studies (see Table 1). Teacher implementation was reported as high in four of these five. We list the specific observation criteria that were used in these studies in Table 2. Only two of these investigators, Levin (1989) and Shortland-Jones (1986), specified their observational criteria. However, even in these two cases, their checklists refer more to the procedures for explicit strategy instruction than to the guided practice of reciprocal teaching.

In nine studies, the experimenter also served as teacher and conducted the lessons, but there was no information on the quality of instruction in eight of the nine. In the remaining study (Palincsar & Brown, 1984), the quality of the instruction was assessed and found to be acceptable, although, as noted, the criteria were not given.

In summary, there was little direct observation of the quality of the reciprocal teaching dialogue in these studies, and no investigator presented a set of criteria specifically designed for evaluating reciprocal teaching. Overall, it was impossible to assess the quality of the actual reciprocal teaching instruction and we were unable to adequately assess whether the instruction in these studies matched the description provided by Palincsar and Brown. But we believe that these studies represent the instruction that occurs in practice as one attempts to implement the reciprocal teaching described by Palincsar and Brown.

This lack of observation and evaluation of instruction is a common problem among studies in which cognitive strategies are taught. The problem is magnified in the case of reciprocal teaching because the dialogue is a critical variable in the instructional procedure.

In later articles, Palincsar and her associates (Palincsar, 1986; Palincsar et al., 1989) illustrated and made explicit the important instructional features of dialogues by comparing the dialogues of several primary teachers. However, none of the studies in this review referenced those articles or referred to those critical features. (We include a list of those features in the Discussion section under the heading “Quality of the Dialogue in Reciprocal Teaching.”)

Assessing student questions and summaries during the dialogues. Another procedure for assessing the quality of the dialogue is to examine the changes in student questions and summaries that occurred during the dialogues. Palincsar and Brown (1984) examined student dialogue and found that the quality of both student questions and student summaries improved significantly from the early sessions to the later sessions. In the early sessions, student questions contained verbatim phrases from the text; in the later sessions, questions were more likely to be paraphrases of main points of text. There was also a shift away from summaries that consisted mainly of details (29% to 4%) and toward summaries

TABLE 2
Teacher implementation

Levin (1989)

There were three full-period observations, by the experimenter, of the control teachers. Four variables were observed:

- (1) students leading the lessons
- (2) students formulating questions about the material
- (3) mention of ISL (Informed Strategies for Learning) strategies
- (4) basal or phonics book in use

For the control classes, no occurrences of variables 1, 2, or 3 were observed; variable 4 was consistently observed.

Experimental teachers were observed four times for 45-minute observations using an ISL Observational Checklist developed by Paris (1984). The checklist included 12 teaching behaviors:

- (1) gains students' attention
- (2) introduces lesson
- (3) states goals
- (4) reviews previous information
- (5) informs students about (a) what it is, (b) how it works, and (c) when and when not to use
- (6) encourages metaphorical reasoning
- (7) persuades why strategy is useful
- (8) models and demonstrates strategy
- (9) allows immediate practice
- (10) provides praise and corrective feedback
- (11) fades support
- (12) bridges to other reading/learning (p. 185)

The experimental teachers had difficulty with behaviors 6, 7, and 8. Although these three behaviors were infrequently observed at the beginning of the program, they were observed most of the time during the end of the program. Remedial work with the teachers was conducted throughout the program. No data were provided on the other nine behaviors found on the ISL checklist.

In addition to observations based on the above ISL checklist, the investigator made three observations of each class during reciprocal teaching sessions and recorded anecdotal field notes. These observations appear to have been focused more on the reciprocal teaching procedure than on the ISL behaviors. The notes revealed two types of teacher error.

Some teachers moved too quickly to take over when students were having problems formulating questions or working out strategies for getting classmates' cooperation (Teachers 1, 3A, 4, and 5). Others (Teachers 2 and 3A) did not interfere even when students made mistakes, but allowed students to mispronounce important words or to accept incorrect answers. (p. 81)

Discussion and review of reciprocal teaching protocols and additional demonstrations of the procedure succeeded in helping most of the teachers overcome the problems mentioned above.

Palincsar (1987)

Two raters, who had conducted reciprocal teaching instruction, listened to an audiotape and rated the quality of instruction. "The raters noted that each of the teachers indicated fidelity to the reciprocal teaching procedure" (p. 18). The specific criteria used to judge the quality were not given in the study.

TABLE 2—continued
Teacher implementation

Palincsar & Brown (1984)

A random selection of audiotapes from five different sessions with each of the three groups allowed an independent evaluator to carry out procedural reliability checks. The evaluator “verified that the procedure was indeed implemented as described” (Palincsar, 1982, p. 74). The specific criteria used to judge the quality were not given in the study.

Shortland-Jones (1986)

The experimenter developed a checklist and observed teachers in four stages of a lesson:

- (1) introduction/motivation
- (2) introduction of the graphic organizer
- (3) students’ generation of metacognitive strategies
- (4) lesson conclusion

“It appears that both groups of teachers [for experimental and control groups] achieved a high rate of implementation of techniques in the introductory stage and in the graphic overview stage” (p. 71).

Taylor & Frye (1992)

Each week, one of the two investigators met individually with each teacher for 5 to 15 minutes to talk about the study, answer concerns, and answer questions. “Based on these discussions it was felt that the teachers were following procedures as requested” (p. 42). No specific rating items were given.

Williamson (1989)

All participating teachers (for both control and experimental groups) were observed by the investigator on three different occasions during regular reading instruction. A checklist containing criteria items related to each of the four reciprocal teaching activities was given to experimental teachers prior to implementation and was used during the observations. The actual checklist or a discussion of the criteria, however, was not provided. Individual conferencing was provided for any experimental teacher having difficulty adhering to the guidelines of reciprocal teaching.

The investigator reported a low degree of implementation overall.

It was observed that not all teachers were equally effective in applying the reciprocal teaching strategy activities. Suggestions and additional coaching did not adequately remedy the situation. (p. 64)

The actual weaknesses found during implementation observations were not discussed in the study.

that included main ideas (52% to 85%). Unfortunately, no other study evaluated the quality of student questions and summaries, not even Palincsar (1987).

Aside from Palincsar’s two studies, only one study (Rich, 1989) contained sample transcripts of complete reciprocal teaching lessons. Mosenthal (1987) developed a procedure for evaluating the adequacy of student questions and summaries during reciprocal teaching dialogues, and we asked him to evaluate two of Rich’s transcripts. He reported coding 90% of the student questions, but only a third of the summaries, as adequate. He also noted that the teacher turned over responsibility for the discussions to the students before they were ready for

the task. Unfortunately, transcripts were not supplied in the other studies (not even in the dissertations, where space was not a problem).

Thus, across these reciprocal teaching studies, the quality of the dialogues was rarely assessed. No criteria checklists were developed that were specific to reciprocal teaching. Few studies provided samples of the actual dialogues. Aside from the original Palincsar and Brown (1984) study, few researchers assessed the quality of student questions and summaries during the dialogues.

Assessing descriptions of reciprocal teaching. Because we were seldom able to assess the quality of instruction directly, we attempted a rough assessment of each investigator's knowledge of reciprocal teaching. We first developed a checklist of criteria based on the description of reciprocal teaching in the Palincsar and Brown (1984) study. (This checklist is presented in Table 3.) We then examined each author's descriptions of the reciprocal teaching research and of his/her own treatment, and recorded the number of checklist items mentioned in those descriptions. (The results are presented in Table 1.) Based on those descriptions, and on our conversations and queries to the researchers, we decided that the quality of instruction met our criteria in 13 of the 16 studies. Of the remaining 3, the investigator noted that implementation may have been poor in 1 (Williamson, 1989) and we were uncertain about the quality of instruction in 2 others (Padron, 1985; Rush & Milburn, 1988).

TABLE 3

Criteria checklist for quality of the description of reciprocal teaching

-
- *1. Students are instructed in a repertoire of strategies (two or more) that they can use to help them better understand what they read.
 - *2. Teacher models each of the activities.
 - *3. Students are invited to make comments regarding the modeling and the passage, such as, "Was there more important information?" or "Does anyone have anything more to add to my prediction?"
 - *4. Students are provided with guided assistance as they participate at whatever level they are capable in carrying out the strategies.
 - *5. Teacher supports each child's participation in the dialogue through specific feedback, praise, prompting, additional modeling, paraphrases, coaching, hints, and explanation.
 - °6. Teacher invites students to initiate discussion and to react to other students' statements. Such participation can include (a) suggesting other questions, (b) elaborating upon a summary, (c) commenting on another's predictions, (d) requesting clarification of material they did not understand, (e) offering additional comments on the content, and (f) helping to resolve misunderstandings.
 - *7. During the reciprocal teaching procedures, there is a gradual shift from the teacher doing much of the work to the child taking over the major thinking role. The teacher gradually transfers control of the dialogues to the students and becomes a supportive observer.
 - °8. During the dialogues, instruction is provided on why, where, and when these strategies might be applied.
-

* This particular characteristic of reciprocal teaching was mentioned in the Palincsar dissertation.

° This particular characteristic of reciprocal teaching was illustrated through the sample dialogues provided in the Palincsar dissertation.

Other Instructional Aids

A few investigators added instruction to reciprocal teaching that might have facilitated or enhanced the outcomes. Rich (1989) instructed her students to make up a question about each paragraph and write a summary before they answered the questions on their daily assessments. If the students were told to apply this strategy to the passages used in the final assessment, it might (or might not) have provided students with additional help.

Both Labercane and Battle (1987) and Williamson (1989) asked the teacher to “think aloud” (Davey, 1983) while modeling the application of the strategies. This modeling of internal thoughts did not appear in the text or dialogues of the original Palincsar (1982) or Palincsar and Brown (1984) reports, but it was included in the description of reciprocal teaching in a later study by Palincsar (1987). Thus, depending on which of Palincsar’s articles they had read, investigators may or may not have included thinking aloud in their instruction. (It is somewhat ironic that neither Labercane and Battle nor Williamson obtained significant results). It is possible that some investigators provided additional aids that had unknown effects on the results.

Grouping Studies by the Quality of the Intervention

All studies included in this review met our selection criteria of comparable control groups and the use of posttests to measure comprehension of new materials. In addition, we divided studies into high-, middle-, and low-quality based on three indicators described above: (a) the use of comprehension probes during the study, (b) assessment of student learning of the strategies, and (c) the quality of the description of the reciprocal teaching dialogue. The ratings are reported in Table 1.

All high-quality studies were high in quality of the dialogues (based on observation or on description in the text), and included both comprehension assessments during the instruction and assessment of the students’ learning of the strategies. Middle-quality studies were also high in the quality of the dialogues (based on observation or on description in the text), but did not include either comprehension assessments during the instruction *or* assessment of strategy learning. Three studies were rated as low or uncertain because their descriptions of reciprocal teaching were lacking in detail, or because the investigator (Williamson, 1989) wrote that the quality of instruction was low. These three also lacked both comprehension probes and assessment of student learning of the strategies. The results, by quality of study, will be discussed further in the following section.

Results

There are four parts to this section. “Major Results” contains the main results on student comprehension. “Results for Instructional Approach by Type of Student by Outcome Measure” is an attempt to look at whether certain students improved in comprehension with either instructional approach or either type of comprehension measure. “Results by Settings” provides results for setting variables such as grade level, size of group, and type of control group. The final part, “Assessing Student Learning of the Strategies,” gives the results for studies that assessed student learning of the cognitive strategies that were taught.

Major Results

This first section contains the major results on student comprehension for the 16 reciprocal teaching studies. These results are given according to (a) type of outcome measure, (b) quality of the studies, (c) number of strategies that were taught, (d) type of instructional approach, (e) types of students that were taught, and (f) researchers conducting the studies.

Type of Outcome Measure

There were three types of outcome measures used in these studies: (a) standardized tests, (b) experimenter-developed short-answer or multiple-choice tests, and (c) experimenter-developed summarization tests. Standardized tests were used in 9 of the 16 studies; experimenter-developed tests were used in 11 of the 16 studies. Both a standardized test and an experimenter-developed comprehension test were used in 5 studies. The results for each study, on each type of outcome measure, are presented in Appendices A and B. The results are grouped according to the three types of outcome measures and expressed as median effect sizes in Table 4.

Of the standardized tests used in nine studies in this review, the Gates-MacGinitie Reading Test was the most frequently cited. Results were significant in only two of these nine studies. The median effect size for studies that used standardized tests was .32.

Experimenter-developed short-answer or multiple-choice tests were used in seven studies, and all but one of those studies (Rich, 1989) used a short-answer format. Students were given 200- to 800-word passages (the median was 425 words) followed by 5 to 10 short-answer or multiple-choice questions. About half the questions were factual questions, and half required inference from the text. Results were significant in six of the seven studies that employed such tests. The median effect size for experimenter-developed short-answer tests was 1.00.

Summarization tests consisted of passages of 250 to 400 words that students were asked to summarize. Various procedures were used to rate the students' summaries, the most common being to score the independent clauses for the level of importance of the ideas they contained. For example, Palincsar (1987) awarded the following points when scoring the summaries: one point for each topic sentence taken from the text, two points if the student invented a topic sentence, one point for each proposition that had been rated by English teachers

TABLE 4
Effect sizes based on quality of study

	Standardized test	Short-answer test	Summarization test	Experimenter- developed test
All studies	.32 (9) ^a	1.00 (7)	.85 (5)	.88 (10)
High quality	.31 (4)	1.00 (3)	.77 (4)	.86 (6)
Middle quality	.36 (3)	1.06 (4)	1.10 (1)	.87 (4)
Low or uncertain quality	-.12 (2)	—	—	—

^aNumber in parentheses refers to number of studies used to compute median effect size.

as “quite important” or “very important” in the text, and additional points for overall clarity and quality of the summaries. In the Taylor and Frye (1989) study, students received one point for each important idea or inference presented or paraphrased in their summaries. Shortland-Jones (1986) scored summaries for the quality and quantity of their ideas. The median effect size for the five studies that used student summaries was .85, and four of the five studies obtained significant results (Table 4).

Overall, the results were quite similar for the short-answer and the summarization tests across these studies. Further evidence of this similarity comes from Palincsar (1987), who gave both summary tests and short-answer tests to her students and obtained similar results for the two types. Rich (1989) asked students to both answer multiple-choice questions and perform a summarization task on the same passage. She obtained similar results for the two types of test. Because of the similarity in results for both types of experimenter-developed comprehension test, we combined the results for the short-answer and the summarization tests into a single measure that we labeled *experimenter-developed comprehension tests*. Overall, then, experimenter-developed comprehension tests were used in 11 studies, and the results were significant in 8 of the 11. The median effect size for these tests was .88 (Table 5). In contrast, the median effect size for the 9 studies that used standardized tests was .32.

Another way of comparing results for standardized tests and experimenter-developed comprehension tests is to look separately at the five studies that used both standardized tests and experimenter-developed tests: Brady (1990), Dermody (1988), Lysynchuk et al. (1990), Taylor and Frye (1992), and Shortland-Jones (1986). In these cases, the pattern is the same. In four of the five studies, the results were nonsignificant on the standardized test but significant on the experimenter-developed test. In the fifth study (Shortland-Jones, 1986), this pattern was reversed. Thus, in studies where investigators used both experimenter-developed outcome measures and standardized tests, the results followed an

TABLE 5
Effect sizes based on quality of study

	Standardized tests	Experimenter-developed tests
All studies	.32 (9) ^a	.88 (10)
High quality	.31 (4)	.86 (6)
Middle quality	.36 (3)	.87 (4)
Low or uncertain quality	-.12 (2)	—
All studies except the three with low or uncertain quality of instruction (Padron, 1985; Rush & Milburn, 1988; and Williamson, 1989)	.36 (7)	.88 (10)

^aNumber in parentheses refers to number of studies used to compute median effect size.

overall pattern: significant findings more frequently on the experimenter-developed tests.

In summary, across the 16 studies, the median effect size was .32 when standardized tests were used and .88 when experimenter-developed comprehension tests were used. Because the results for the standardized tests and the experimenter-developed comprehension tests were quite different, even when both measures were used in the same study, we kept the two sets of results separate in all our analyses and did not provide any overall effect sizes across both types of test.

Results by Quality of Study

We made a number of groupings according to our assessment of the quality of the studies. In Table 4, the effect sizes are grouped by the three major outcome measures: standardized tests, experimenter-developed short-answer and multiple-choice tests (usually short-answer questions), and experimenter-developed summarization tests. Table 4 shows that there was little difference between the results for the studies classified as “high-quality” and those classified as “middle-quality” on the three outcome measures. For both the high-quality and the middle-quality studies, moderate effect sizes were obtained when a standardized test was used and much higher effect sizes were obtained when experimenter-developed comprehension tests, either short-answer or summarization, were used. This pattern remained the same when we combined results for the short-answer tests and the summarization tests.

The studies that we classified as low-quality did not use any experimenter-developed comprehension tests. Excluding the three studies of low or uncertain quality did not have any effect upon the overall results (Table 5).

We decided to retain the three studies of low or uncertain quality in our analyses for two reasons. First, we were not sure we had identified all of the studies that were low on implementation. Only Williamson (1989) explicitly said that there was low implementation in her study, but she may have said this as a way of explaining why she did not obtain significant results. There may have been studies with significant results that were also low in implementation, but where the investigator may have been unaware of it and/or chose not to discuss it. A second reason for keeping all 16 studies was that these studies, as a group, represent reciprocal teaching “in use.” They represent reciprocal teaching as implemented by people concerned with doing a good job—doctoral students and professors who had to submit their work for review. They represent what we might expect when people take the Palincsar and Brown articles and try to implement reciprocal teaching in schools. Therefore, we decided to include the “low” and “uncertain” studies in our analyses because the larger set of studies would provide a broader picture of the results obtained when reciprocal teaching is implemented in classrooms.

Number of Strategies Taught

Palincsar and Brown (1984) taught four specific comprehension-fostering strategies during reciprocal teaching: questioning, summarizing, predicting, and clarifying unknown words. In 12 of the 16 studies, teachers provided instruction on these four strategies. The remaining 4 studies taught 2, 3, or 10 cognitive strate-

gies. We were interested in whether there were differences in results according to the number of strategies taught. In Table 6, the results for the studies that provided instruction in the 4 strategies are compared with the results for the other studies, which taught 2, 3, or 10 strategies.

For standardized tests, the eight studies that taught 4 strategies had a median effect size of .20. The single study that used a standardized test and taught a different number of strategies, Shortland-Jones (1986), obtained an effect size of .77. For experimenter-developed comprehension tests, the median effect size for the seven studies that taught the 4 strategies was .85, compared with 1.24 for the three studies that taught a different number of strategies. (As noted earlier, five studies used both a standardized test and an experimenter-developed comprehension test). No relationship was found between the number of strategies taught and student achievement; studies obtained significant results both when the investigators taught the 4 strategies identified by Palincsar and Brown, and when they used 2, 3, or 10 strategies.

Type of Instructional Approach

The reciprocal teaching dialogue was the main instructional vehicle in all the studies. However, as noted earlier, two variations of reciprocal teaching were

TABLE 6
Effect sizes for overall results by type of test

	Standardized tests	Experimenter-developed tests
All reciprocal teaching studies	.32 (9) ^a	.88 (10)
All studies except the 3 with uncertain or low quality of instruction (Padron, 1985; Rush & Milburn, 1988; and Williamson, 1989)	.36 (7)	.88 (10)
All reciprocal teaching studies providing instruction on 4 strategies	.20 (8)	.85 (7)
All reciprocal teaching studies providing instruction on 2, 3, or 10 strategies	.77 (1)	1.24 (3)
Studies that taught 4 strategies to only good/poor students	.29 (2)	.88 (3)
Palincsar/Brown studies	—	.94 (2)
Other studies	.29 (2)	.68 (1)
All studies except Palincsar & Brown (1984) and Palincsar (1987)	.32 (9)	.86 (8)
All studies except Palincsar & Brown (1984), Palincsar (1987) and the 3 studies with uncertain or low quality of instruction (Padron, 1985; Rush & Milburn, 1988; and Williamson, 1989)	.36 (7)	.86 (8)

^aNumber in parentheses refers to number of studies used to compute median effect size.

used in these studies. In one approach, 4 to 20 days of explicit teaching of cognitive strategies took place *before* the dialogues began (ET-RT). In the second approach, all instruction took place *during* the reciprocal teaching dialogue (RTO). We were interested in whether these two variations yielded different results.

When standardized tests were the outcome measure, almost all results were nonsignificant, and so there were no differences between the two approaches.

There were eight studies that used ET-RT and experimenter-developed comprehension tests. Results were significant in six of those eight studies, with a median effect size of .86. There were three that used the RTO approach and experimenter-developed comprehension tests. Two of the three obtained significant results with a median effect size of 1.0. So there was little difference between the two instructional approaches, as assessed by either standardized tests or experimenter-developed comprehension tests. These issues are explored further in the next section.

Type of Student

We divided studies into three types: (a) those in which all the students in a given class or classes were used, (b) those in which only students who were good in decoding but poor in comprehension (*good-poor*) were selected, and (c) those studies in which students classified as poor readers were used, but where no attempt was made to determine their decoding ability. We labeled the third group *below average*. The results for studies that taught each type of student are presented in Table 7.

There was no difference in results between studies in which all students in a classroom were used and studies in which good-poor students were used. The median effect sizes for both groups were about .30 for standardized tests, and .86 for experimenter-developed comprehension tests (Table 7). Therefore, we combined the studies from these two groups into a single category that we labeled *good-poor or all*.

In the two studies by Palincsar (Palincsar & Brown, 1984; Palincsar, 1987), only good-poor students were selected. We compared results for the Palincsar (1987) and Palincsar and Brown (1984) studies with the results for the two other studies that also selected good-poor students, and the results were similar for these four studies.

For below-average students, the median effect size for standardized tests was somewhat lower than when good-poor or all students were taught (median effect size = .08) (Table 7). The effect size for experimenter-developed comprehension tests was somewhat higher, but this was based on only two studies. We kept the

TABLE 7
Effect sizes for type of student \times type of test

	Standardized tests	Experimenter-developed tests
All students	.32 (4) ^a	.85 (5)
Good-poor students	.29 (2)	.88 (3)
Below-average students	.08 (4)	1.15 (2)

^aNumber in parentheses refers to number of studies used to compute median effect size.

studies that used below-average students separate from studies that used good-poor or all students in many of our analyses due to interest in these students.

Other Researchers

We were also interested in how well reciprocal teaching fared in the hands of other researchers, and so we excluded the two studies by Palincsar and Brown (1984) and Palincsar (1987) and recomputed the median effect sizes (Table 6). Overall, there was little change in the results. In another analysis, we excluded the two studies by Palincsar and Brown (1984) and Palincsar (1987) and also excluded the three studies of low or uncertain quality (Table 6). None of the three studies of low or uncertain quality used experimenter-developed comprehension tests, and so excluding those three studies had no effect upon the results for experimenter-developed comprehension tests.

Summary of Major Results

Across all the studies, the median effect size was .32 when standardized tests were used, and .88 when experimenter-developed comprehension tests were used. These results remained the same when we removed the three studies that we labeled as low or uncertain in quality. Results were generally the same whether students were taught the 4 cognitive strategies used by Palincsar and Brown, or 2, 3, or 10 cognitive strategies. Results were also the same when we excluded the two studies by Palincsar (1987) and Palincsar and Brown (1984). Results were generally the same when all students in a classroom or classrooms were used as when only good-poor students were used. When poor readers were selected without regard to their decoding ability, the effect sizes were somewhat higher when experimenter-developed comprehension tests were used and somewhat lower when standardized tests were used.

Results for Instructional Approach by Type of Student by Outcome Measure

In the previous section, we identified two types of outcome measures (standardized tests and experimenter-developed comprehension tests), two groupings of students (good-poor/all and below-average), and two overlapping instructional approaches (teaching of strategies both before and during the dialogues [ET-RT] and teaching strategies only during the dialogues [RTO]). We were interested in any patterns across these eight cells. The results are presented in Table 8. Because of the small number of studies in most of the cells, effect sizes are not given; rather, results are presented as ratios of significant results to the number of studies.

Table 8 shows that when experimenter-developed comprehension tests were used, the results were almost always significant, regardless of the type of student

TABLE 8
Ratio of significant studies to total studies

Type of outcome measure	Good-poor/all students		Below-average students		Overall
	ET-RT	RTO	ET-RT	RTO	
Standardized tests	1/5	0/1	1/2	0/3	2/11
Experimenter-developed tests	5/6	1/2	1/1	1/1	8/10

or instructional approach. The most common method was to use the ET-RT approach with good-poor or all students. When experimenter-developed comprehension tests were used, this modal approach had a median effect size of .85.

Table 8 also shows that results were seldom significant when standardized tests were used, regardless of type of student or instructional approach. The same combination—the ET-RT approach with good-poor or all students—that was quite effective when experimenter-developed comprehension tests were used was seldom significant when standardized tests were used.

Test by Type of Student

There were two patterns that emerged when we looked at results by type of student (good-poor/all or below-average). First, experimenter-developed tests were more often used in studies that taught good-poor or all students, and these tests usually produced significant results. Second, standardized tests were used proportionally more often with below-average students, and these tests usually yielded nonsignificant results.

Instructional Approach

As noted in the introduction, Palincsar has used two related approaches for teaching the cognitive strategies. The first was RTO, in which all strategy instruction took place during the dialogues (Palincsar & Brown, 1984). In the second approach, the teachers provided three to six days of “direct instruction in each strategy” before the dialogues began (Brown & Palincsar, 1989, p. 33; Palincsar et al., 1990, p. 15). This second approach was used in Palincsar (1987). Both approaches were also found in the other 14 reciprocal teaching studies. We were interested in whether the two approaches yielded different results. The results for these two approaches are presented in Table 8. Effect sizes are not given because of the small number of studies in these eight categories.

There were six studies in which good-poor or all students were taught, experimenter-developed comprehension tests were used, and the ET-RT approach was used. Five of these six obtained significant results. Unfortunately, we did not find enough instances in the remaining three categories to provide any contrast.

Results by Settings

We also investigated any pattern in the results by grade level, number of instructional sessions, the size of the instructional group, the person providing the instruction, and the type of control group that was used.

Grade Level

Instruction in reciprocal teaching was provided for students over a wide range of grade levels, from seven-year-olds to adults. When we looked at results by grade level (see Table 9), we found both significant and nonsignificant effects at all grade levels, with the exception of third grade. Three third-grade classes in three separate studies (two in the good-poor or all sample, and one in the below-average sample) yielded nonsignificant results. However, in two of these three studies, the investigator used standardized tests. So there were significant results from grades four through adult, and insufficient data to determine the effectiveness of reciprocal teaching with third-grade students.

TABLE 9
Grade level of students

Good-poor/all			Below-average		
Significant	Mixed	Not significant	Significant	Mixed	Not significant
6-8	4 & 7	3	4-6	7	3-5
4 & 6	5-6	3	adult		5
7	4	Vocational			8-12 yr. olds
	1-6	College			

Number of Instructional Sessions

The number of instructional sessions provided in these studies ranged from 6 to 100 (see Appendixes A and B). We found no difference in the median number of sessions for significant, mixed, and nonsignificant studies across different types of students (see Table 10). Two of the studies with the largest numbers of sessions (50 and 100) had nonsignificant results, but those studies used standardized tests, and, as was noted earlier, 8 of the 10 times standardized tests were used, the results were nonsignificant. When we excluded studies that used only standardized tests, the range was from 6 to 25 sessions, and no relationship was found between number of sessions and significance of the results. No differences in the number of sessions were noted when we compared studies that used ET-RT with those that used RTO.

Size of Instructional Group

Table 11 presents results by size of the instructional group. We found no relationship between the size of the instructional group and the significance of the results for either type of student. Even after eliminating the studies that used only standardized tests, the range of group size for studies with significant results was from 2 to 23 students.

Person Providing the Instruction

We were interested in whether different results occurred when the experimenter or a classroom teacher provided the instruction. These results are summarized in Table 12. We detected no differences. Studies were equally effective whether an experimenter or a teacher provided the instruction.

TABLE 10
Number of instructional sessions

Good-poor/all			Below-average		
Significant	Mixed	Not significant	Significant	Mixed	Not Significant
25	13	50	6	25	100
20	11	20	50		28
20	24	9			8
	20				

TABLE 11
Size of instructional group

Good-poor/all			Below-average		
Significant	Mixed	Not significant	Significant	Mixed	Not significant
12	4	14	18	6	4
18	23	5	10		4
2	7	4			22
	13				

Types of Control Groups

There were four types of control groups used in these studies: *basal*, *basal with additional instruction*, *read and answer questions*, and *other activities*. An explanation of each type of control group follows. We were interested in knowing if different types of control groups were associated with different results.

Basal. This category refers to those studies in which the control group continued the traditional practice of reading stories and answering questions, and received no further instruction on how to construct answers to these questions. Control groups from 5 of the 16 studies were placed in this category.

Additional instruction. We placed studies in this category when the control group received explicit instruction in reading comprehension, such as instruction in identifying the main idea of a paragraph (Lonberger, 1988), or in answering questions whose answers were explicitly or implicitly present in the text (Padron, 1985; Palincsar & Brown, 1984). The control groups from 6 of the 16 studies were placed in this category.

Read and answer questions. This category includes studies in which the control group read silently the same training passages used by the students in the experimental group. Two studies fit this description. In one study (Rich, 1989), the students answered questions on the passages. Students in the second study (Lysynchuk et al., 1990) read the training passages, but we could not determine from the text whether those students also answered questions.

Other. We placed in this category those studies that did not fit into any of the other categories. There were three such studies: one in which the control group did computer extension exercises not directly related to reading comprehension but did take the daily comprehension tests (Brady, 1990), and two studies in which we were not told what the control group did (Dermody, 1988; Labercane & Battle, 1987).

TABLE 12
Person providing instruction

	Significant	Mixed	Not significant
Experimenter	2	3	2
Teacher	3	2	3
Student leaders	—	—	1

One might argue that the first two types of control group—basal and additional instruction—were the most appropriate because they were most closely related to traditional instruction. Accordingly, we recomputed the overall results just for those 11 studies that used the first two types of control groups—basal and additional instruction. Separate results are presented for experimenter-developed comprehension tests and standardized tests (see Table 13). The effect size for those 11 studies was the same as the overall effect size that was obtained for all studies that used experimenter-developed comprehension tests. The results were slightly lower for studies with the first two types of control group when standardized tests were used. When we excluded from these 11 the 3 studies of low or uncertain quality, the results remained the same when experimenter-developed comprehension tests were used, and were lower still when standardized tests were used.

Summary

When we compared studies that had significant results with studies that had nonsignificant results, we found no differences in results by (a) grade level, (b) number of sessions, (c) size of the instructional group, (d) number of cognitive strategies that were taught, or (e) whether the investigator or the teacher did the training. There was no decrease in the overall effect size for experimenter-developed comprehension tests when we used only studies that had more traditional control groups, but the effect size for those studies using standardized tests was a bit lower in that context.

Assessing Student Learning of the Strategies

In 8 of the studies, as noted earlier, the investigators assessed the extent to which the students had learned the specific strategies. Four different assessments were used. The studies that used these assessments are listed in Table 14, and the results are discussed in this section.

Asking Students to Generate Questions About a New Passage

In the Palincsar and Brown (1984) study, at the completion of the reciprocal teaching sessions, the investigators assessed student ability to generate questions.

TABLE 13
Effect size for studies providing reading instruction to control groups

	Standardized tests	Experimenter-developed tests
Studies providing reading instruction to control groups	.07 (5)	.87 (6)
Studies providing reading instruction to control groups with the exception of the 3 low- or uncertain-quality studies (Dermody, Rush & Milburn, Williamson)	-.14 (4)	.87 (6)

TABLE 14
Assessing student learning of the strategies

Asking students to generate questions about a new passage

Dermody, 1988
 Lysynchuk et al., 1990
 Palincsar, 1987
 Palincsar & Brown, 1984
 Shortland-Jones, 1986
 Taylor & Frye, 1992

Asking students to summarize a passage

Dermody, 1988
 Lysynchuk et al., 1990
 Palincsar, 1987
 Palincsar & Brown, 1984
 Shortland-Jones, 1986
 Taylor & Frye, 1992

Asking students to make predictions

Dermody, 1988

Asking students how and/or why they would use strategy

Levin, 1989
 Lonberger, 1988
 Padron, 1985

Students were given two passages (written at their seventh-grade level) and asked to “write 10 questions a classroom teacher might ask if testing a student’s knowledge of a passage” (p. 134). The questions were scored by two independent raters, who

were asked to rate each question as a main idea question (worth two points) or a detail question (worth one point), as a question lifted directly from the text (zero points) or paraphrased (one point). In addition, the overall quality of each question was rated on a five point scale from one to five. Finally, a question that the rater indicated she would ask herself was awarded an extra point. (p. 150)

We present this assessment procedure in some detail because it is sensible, and it is one that might well be used in future research.

Surprisingly, although question generation was one of the strategies taught in all of the studies, the ability to generate questions was assessed in only six studies: Dermody (1988), Lysynchuk et al. (1990), Palincsar (1987), Palincsar and Brown (1984), Shortland-Jones (1986), and Taylor and Frye (1992). In each of the six studies, students in the reciprocal teaching group were superior to control students on either an experimenter-developed comprehension test or a standardized test. However, in five of these same six studies, there was no difference between the reciprocal teaching group and the control group in the level of questions generated (Lysynchuk et al., 1990; Palincsar, 1987; Palincsar & Brown, 1984), the number of questions (Taylor & Frye, 1992), or a rating of the quality of the questions (Shortland-Jones, 1986). In other words, we found no relationship between posttest measures of ability to generate questions and reading comprehension scores.

Coding Student Questions and Summaries

As noted earlier, changes in the quality of student questions and summaries during the dialogues was assessed in only one study, the original Palincsar and Brown (1984) study. They found that the quality of both student questions and summaries improved significantly from the early sessions to the later sessions. In the early sessions the student questions contained verbatim phrases from the text; in the later sessions the questions were more likely to be paraphrases of main points of text. There was also a shift away from summaries that consisted mainly of details (29% to 4%) and toward summaries that included main ideas (52% to 85%). Unfortunately, no other study evaluated the quality of the student questions and summaries during the dialogues.

Asking Students to Summarize a Passage

In six of the studies, after the instruction was completed, students were asked to read a new passage and to write a summary of it (Dermody, 1988; Lysynchuk et al., 1990; Palincsar, 1987; Palincsar & Brown, 1984; Shortland-Jones, 1986; Taylor & Frye, 1992). The summary was then rated, usually by scoring each idea for its importance in the passage. Significant differences favoring the experimental group were obtained in four studies (Dermody, 1988; Palincsar, 1987; Palincsar & Brown, 1984; Taylor & Frye, 1992) but not in the fifth (Shortland-Jones, 1986). In the sixth study (Lysynchuk et al., 1990), one grade had significant results and one grade did not.

We suggest that a measure of student ability to summarize is both a measure of implementation and also a legitimate comprehension outcome measure; therefore, we also used summarization as an outcome measure in the previous section.

Asking Students to Make Predictions

Although students were taught to make predictions in 12 of the 18 studies, they were assessed on this ability in only one study. Dermody (1988) asked students to make predictions on what a story might be about by simply reading the title of the story. The results significantly favored the experimental group over the control group on ability to make predictions. (Concerns regarding this prediction strategy, as expressed by Brady, are discussed under the heading "Which of the Major Reciprocal Teaching Strategies Were Most Helpful?")

Summary

Four procedures were used to assess student learning of the strategies, but only two procedures appeared in more than one study: student ability to generate questions and student ability to summarize a portion of text. Interestingly, in most of the studies there was no difference between experimental and control students in ability to generate questions; nonetheless, the experimental students were usually superior on the comprehension measure. As will be discussed later, we interpret this result as suggesting that the students in the experimental group learned some additional processing procedures that were not assessed by simply asking students to generate questions.

Discussion

Summary of All Results

Reciprocal teaching overall. When standardized tests were used, the reciprocal teaching treatment was significantly superior to the control treatment in 2 of 11 studies, with a median effect size (*es*) of .32. When experimenter-developed comprehension tests were used, students in the reciprocal teaching treatment had scores that were significantly superior to those of the control group in 8 of 10 studies, with a median *es* of .88.

Type of experimenter-developed comprehension tests. Two types of experimenter-developed comprehension tests were used in these studies: short-answer tests, and tests that asked students to summarize passages. The results were similar for both types of test, and these results were combined into a category labeled *experimenter-developed comprehension tests*.

Quality of the studies. When separate analyses were made for studies classified as “high” and “middle” in quality, the results were similar on both experimenter-developed comprehension tests and standardized tests. The three studies classified as “low” or “uncertain” in quality used only standardized tests, so the overall results for experimenter-developed comprehension tests are unaffected by those studies. The median standardized test’s effect size for studies that we classified as “low” or “uncertain” in quality was smaller than the median standardized test’s effect size for studies classified as either “high” or “middle” in quality.

Instructional approach. When (a) there was explicit teaching of cognitive strategies before beginning the reciprocal teaching dialogue, and (b) results were assessed using experimenter-developed comprehension tests, then results were significant in six of seven studies with a median effect size of .88. Only three studies used the reciprocal-teaching-only approach together with experimenter-developed comprehension tests: the results were significant in two of those three studies.

Type of student. Overall, the results were quite similar for studies that used all students in a classroom and for studies that used only students who were good in decoding ability but poor in comprehension skills (*good-poor*). For both types of students, a median effect size of .30 resulted when standardized tests were used and a median effect size of .85 was obtained when experimenter-developed comprehension tests were used. Only two studies with below-average students used experimenter-developed comprehension tests, and both of these obtained significant results.

Instructional approach by type of test by type of student. Reciprocal teaching was most effective when experimenter-developed tests were used. These results were strongest when experimenter-developed comprehension tests were used in combination with (a) explicit teaching of strategies before the reciprocal teaching dialogues began (ET-RT) and (b) good-poor or all students. The results were particularly weak when standardized reading tests were used in combination with (a) strategy instruction given *during* the reciprocal-teaching dialogues only (RTO) and (b) below-average students.

Good-poor students and four strategies. There were only two additional studies that used the original Palincsar and Brown (1984) design (four cognitive strategies taught to good-poor students, using experimenter-developed comprehension

tests). These were Lysynchuk et al. (1990) and Palincsar (1987). Their results were similar to the those obtained by Palincsar and Brown (1984).

Grade level of students. When experimenter-developed comprehension tests were used, significant results were obtained in grade four through adult, with no pattern favoring any particular grade.

Number of instructional sessions. When studies that used experimenter-developed comprehension tests were considered separately, the number of instructional sessions ranged from 6 to 25. No relationship was found between the number of sessions and significance of the results. Two studies had large numbers of sessions (50 and 100), but they used standardized tests and obtained nonsignificant results.

Size of instructional group. Instructional group size ranged from 2 to 23 students. We found no relationship between the size of the instructional group and the significance of the results.

Number of strategies taught. In 12 of the 16 studies, the strategies taught were the four cognitive strategies first identified by Palincsar and Brown (1984). The remaining four studies, which taught 2, 3, or 10 strategies, were as effective as those that taught the 4 original strategies.

Experimenter versus teacher. Results were equally effective whether the experimenter or a teacher provided the instruction.

Type of control group. When we considered only those 11 studies in which the control group received basal or additional reading instruction, the results for the studies that used experimenter-developed comprehension tests were the same as the results for the remaining studies. However, the median effect size for the studies that provided basal or additional reading instruction to the control group and used standardized tests dropped to .07, and dropped further to $-.14$ when the three studies of low or uncertain quality were eliminated (Table 13).

The Contributions of Palincsar and Brown

We would be remiss if we did not begin this discussion by acknowledging four contributions of Palincsar and Brown (1984) to the teaching of reading comprehension. The first contribution is their introduction of the terms *comprehension-fostering* and *comprehension-monitoring*. Their focus on teaching students comprehension-fostering procedures builds on the earlier work of Durkin (1979), who noted that most teachers she observed asked questions about a story and did not teach comprehension strategies. Palincsar and Brown's work shifts the role of the teacher to that of helping students acquire specific comprehension-fostering and comprehension-monitoring activities that the students can use to monitor and facilitate their own comprehension.

A second contribution is the identification of four specific comprehension-fostering and comprehension-monitoring activities: self-questioning, summarizing, clarifying, and predicting. Which strategies are most productive is an open question that we will discuss below, but their selection of only four strategies is a great improvement over the previous practice of listing over 150 "reading skills" in basal reader workbooks.

A third contribution is their emphasis on practicing these strategies within the context of reading actual materials, in contrast to the more common method of practicing these strategies only when completing worksheets. Within this practice

they provided scaffolds to support student learning and expanded the idea of students providing support for one another.

A fourth contribution is Palincsar and Brown's refinement and popularization of the instructional concept of scaffolding. This term, and the different instructional procedures it represents, help provide a vocabulary and an instructional approach to teaching difficult-to-learn tasks.

Standardized Tests and Experimenter-Developed Comprehension Tests

Most of the results in these studies were significant when experimenter-developed comprehension tests were used (median effect size = .88). However, few of the results were significant when standardized tests were used (median effect size = .32). The results obtained when experimenter-developed comprehension tests were used were generally the same regardless of the number of strategies that were taught, whether one included or excluded the studies by Palincsar and Brown, and whether one included or excluded the studies of low or uncertain quality. All experimenter-developed comprehension tests in these studies assessed student comprehension of *new* materials—that is, materials that were not used in the reciprocal teaching lessons.

The pattern of obtaining higher effect sizes when experimenter-developed comprehension tests were used was also found by Slavin (1987) in his review of studies that used mastery learning. Slavin found a higher median effect size when experimenter-developed comprehension tests were used to measure student achievement (median *es* = .24) than when standardized tests were used (median *es* = .04). The mastery learning studies included social studies, science, and mathematics instruction. Slavin's summary of results suggests that this pattern, favoring experimenter-developed comprehension tests, may extend beyond reading comprehension.

These discrepancies, between the experimenter-developed comprehension test results and the standardized test results, led us to compare one of the standardized tests and one of the experimenter-developed comprehension tests. We selected the Gates-MacGinitie reading test, Second Edition (MacGinitie, 1978), because it was the most frequently used standardized test in these studies. Level D of that test (Grades 4, 5, and 6) was used in four studies (Brady, 1990; Labercane & Battle, 1987; Lysynchuk et al., 1990; Taylor & Frye, 1992). We also looked at the passages and questions developed by Palincsar (1982) because these passages had been used in three studies (Palincsar & Brown, 1984; Palincsar, 1987; Brady, 1990).

The comprehension section of Level D of the Gates-MacGinitie test contains 16 paragraphs. These paragraphs are 44 to 144 words long, with a median of 84 words. Most of the paragraphs were taken directly from expository and narrative literature, or adapted from that material. Three paragraphs were composed especially for the test. Each paragraph is followed by two to four multiple-choice questions. Almost all of the questions require varying degrees of inference to answer.

The experimenter-developed comprehension tests, used in the reciprocal teaching studies, consisted of a 200- to 900-word passage that the students had not seen before (the median was 400 words), usually followed by 8 to 10 short-answer or multiple-choice questions. The passages used by Palincsar consisted

of expository material taken from the reading series published by Scott Foresman and by Ginn and Company. These tests had a mixture of short-answer factual questions and questions that required integrating parts of the text.

The Gates-MacGinitie test and Palincsar's passages differ in a number of dimensions. These differences include the length of the passages, use of topic sentences, the amount of search required to answer a question, the amount of conceptual knowledge required to answer the inferential questions, the vocabulary, and the breadth of the sources.

As noted, the passages in the experimenter-developed comprehension tests were longer. It is possible that answering questions from these longer passages was easier for students because they could use the larger context to help them. The limited context of the Gates-MacGinitie paragraphs may have made answering those questions more difficult.

The passages also differed in their use of topic sentences. Only 1 of the 16 paragraphs used in Level D of the Gates-MacGinitie test followed the topic-sentence-and-supporting-detail format. There were paragraphs that described greenhouses and porcupines that did not contain topic sentences. Thus, in order to answer questions about the topic of a paragraph, a student had to read first to identify a main topic, and then reread to pull details together.

In contrast, the 400- to 450-word passages in the experimenter-developed comprehension tests were reader-friendly. Typically, the first paragraph gave an overview of the topic and the remaining paragraphs elaborated on the opening paragraph. Almost all these paragraphs began with topic sentences and then provided supporting details. For example, a paragraph began with "Some conservationists specialize in water problems," and the remainder of the paragraph provided details on this work. Each paragraph in a passage about minerals and their locations, compositions, and atoms began with a topic sentence.

The amount of search and rereading needed to answer a question was greater in the Gates-MacGinitie test. For example, a paragraph of 83 words about the city of Carcassonne gives so many details about tractors, castles, walls, knights, and military parades that a student had to reread it, more than once, to determine which detail was most important. An adapted paragraph begins with two sentences that contain many details about a dog's "language," and then shifts to three sentences on how a dog understands humans. A question then restates four details and asks the student to select the main one. Rereading was almost always necessary. In contrast, the passages that Palincsar selected seldom contained these problems; rereading was seldom necessary when we tried to answer the questions on those passages.

A fourth difference is that the inferential questions in the Gates-MacGinitie test required greater conceptual knowledge. The story on greenhouses required a student to infer, from the description of the greenhouse, what would happen to the earth's surface if there were more carbon dioxide and water vapor around the earth. The inferential questions on the experimenter-selected passages were never so complex; they usually required combining no more than two adjacent sentences. For instance, in order to answer a question on why a cowboy's boots were made a special way, a student had merely to combine the sentence that said, "There were two unusual characteristics about a cowboy's boots," with the next sentence that explained the reasons for the high heels and the pointed toes.

The vocabulary was also more complex in the Gates-MacGinitie test. For example, one passage explained that the source of the light in a firefly fish is “a colony of bacteria that live there, nourished by secretions from the fish.” To answer one of the following questions, a student had to know that “nourished by secretions” refers to how the fish provides food for the bacteria. There is no way to deduce the meaning of “nourished by secretions” from context in this 109-word paragraph.

The Gates-MacGinitie paragraphs came from a wider variety of sources and were excerpted from books and magazine articles. The experimenter-selected passages came from material that had been written or rewritten specifically for students in elementary and intermediate grades.

In summary, the experimenter-developed comprehension tests seemed easier to answer because (a) the passages were longer, (b) the passages were almost always organized in a topic-sentence-and-supporting-detail format, and (c) answering the questions required less background knowledge and less searching of the text. In addition, the passages on the experimenter-developed tests were similar to those practiced during the reciprocal teaching dialogue. In contrast, many of the paragraphs in the Gates-MacGinitie test came from a wider variety of literature, which was not discussed during reciprocal teaching. We do not know what results would have been obtained if passages from the type of literature used in the Gates-MacGinitie test had been studied during reciprocal teaching sessions.

The Role Of Cognitive Strategies

Reciprocal teaching is part of the larger body of research on the teaching of cognitive strategies. Three interesting questions concerning the general topic of cognitive strategy instruction appeared in these studies: What are the cognitive changes that occur when strategies are taught; how many and which strategies should be taught; and which of the major reciprocal teaching strategies were most helpful to students?

What Cognitive Changes Occur When Strategies Are Taught?

As discussed in the Results, under “Assessing Student Learning of the Strategies,” there were six studies in which, after the instruction, students were tested on their ability to generate questions. In all six studies, students in reciprocal teaching groups were superior to control students in comprehension skills, whether evaluated by experimenter-developed comprehension tests or standardized tests. However, in five of these same six studies, there was no difference between the reciprocal teaching groups and the control groups in the level of questions generated (Lysynchuk et al., 1990; Palincsar, 1987; Palincsar & Brown, 1984), the number of questions (Taylor & Frye, 1992), or a rating of the quality of the questions (Shortland-Jones, 1986). In other words, we found no relationship between posttest measures of ability to generate questions and reading comprehension scores.

This same pattern also occurred in two other studies that involved student questioning. In a study conducted by Risko and Feldman (1986), no relationship was found between the type of question that three remedial second-grade students asked about a story and their ability to answer that type of question on a posttest.

In a study on reciprocal questioning by Manzo (1969), students practiced generating and answering questions, but without receiving any training in asking questions. During the study, the quality of the questions asked by students in the experimental group improved so that, near the end of the study, most of the questions generated by the students in the experimental group were questions requiring translation, conjecture, and explanation. In contrast, few questions asked by students in the control group were of these higher-level types. However, despite these differences in questioning, there were no significant differences between experimental and control students' scores on standardized reading tests. Thus, in these two additional studies, there was, again, no correlation between the ability to generate questions and performance on a comprehension measure.

These findings—the lack of a correlation between the ability to generate questions and student comprehension of new text—reinforce Resnick's (1987) contention that strategies such as question generation and summarization may not be the actual strategies used by skilled performers. Resnick stated:

Research has located a "psychological space" in which educationally powerful effects seem to occur, but it has not yet adequately explained what happens in that space to produce the effects. (p. 27)

We suggest, albeit tentatively, that in the successful studies that taught question generation, what the students learned was not simply how to generate questions. Rather, as explained by Palincsar and Brown (1984), the new strategies enabled and required the students to perform deeper processing of what they read, to engage in making sense of what they read, to be aware of when they did not understand the material, and to engage in additional reading and searching when they encountered comprehension difficulties.

Brady also suggests that the reciprocal teaching practice may have changed the way students read in ways beyond the specific strategies that were taught.

The continual emphasis on making meaning that is required by the questioning, summarizing, and predicting activities actually forced the students to move beyond a belief that decoding the words was a sufficient response to a request that they 'read' a passage. This emphasis on making meaning pushed them through the decoding and into making sense of what they read. I felt that for some of them, this was the first time the concept of reading included this element as a major concern. (personal communication, September 26, 1991)

Brady is suggesting that, for students in his study, the process of trying to make sense of what they read was more important than the specific strategies they were taught.

In other words, there may be a difference between the cognitive strategy that was taught and the strategy that was practiced and learned. However, it may also be possible that practicing the strategies facilitated the development of new processing strategies. There is a need for more research on changes in student processing before and after this type of instruction.

Resnick (1987) notes that if we cannot produce a more substantial explanation of the internal events that produce the improved comprehension, it will be difficult to develop a training approach (p. 27). However, work by Bereiter and Bird (1985) suggests that it is possible to develop effective cognitive strategies even

in the absence of theory. Bereiter and Bird gave passages that contained ambiguous sentences to good readers and asked them to think aloud as they read. Bereiter and Bird studied the protocols of these readers and identified four comprehension strategies that they used: restating material, rereading, demanding relationships between text information, and formulating the difficulties as problems. Bereiter and Bird then explicitly modeled and taught these same strategies to less able readers and demonstrated that teaching such strategies can lead to improved comprehension. Such an approach might be used by others to develop and teach cognitive strategies in the absence of theory.

This “think aloud” approach might also be used to assess changes in students’ processing before and after cognitive strategy instruction.

How Many and Which Strategies to Teach?

In these 16 reciprocal teaching studies, investigators taught from 2 to 10 cognitive strategies. Most of the studies taught the 4 strategies introduced by Palincsar and Brown (1984): question generation, summarization, clarification, and prediction. However, no relationship existed between the number of strategies taught and the results (see Table 12). Studies that taught 2 strategies, 4 strategies, or 10 cognitive strategies all produced significant gains in comprehension. Although all studies included the 2 strategies of question generation and summarization, there were differences across studies in which and how many additional strategies were taught, and few differences in effects were noted.

There are a number of studies outside the reciprocal teaching tradition that taught only a single strategy and also obtained significant results. Consider three reciprocal-teaching-like studies that were precursors to the Palincsar and Brown study (Helfeldt & Lalik, 1976; Manzo, 1969; Nolte & Singer, 1985). In these studies, only one strategy (question generation) was taught, and the instruction did not contain the expert scaffolding and gradual withdrawal of support that characterizes reciprocal teaching. Nonetheless, two of these three studies yielded significant results. Other studies outside the reciprocal teaching tradition that taught only the single strategy of question generation also obtained significant results (Blaha, 1979; Cohen, 1983; Davey & McBride, 1986).

At present, one cannot make a definitive statement on which strategies and combinations of strategies are most effective. There is a need for more research on the effects of teaching different individual strategies and combinations of strategies. If comprehension can be improved with a single strategy, why teach four?

Which of the Major Reciprocal Teaching Strategies Were Most Helpful?

Only one investigator (Brady, 1990) discussed which of the four major reciprocal teaching strategies were most helpful to children. He reported that two of the strategies, clarification and prediction, were difficult to use when studying history text. Brady wrote that when a student asks for clarification of a term, the teacher and the other students normally use the surrounding context to help determine the meaning. However, Brady noted that, in the text he was using, the writing was so dense that students were seldom able to determine the meaning of a word by using context, or the glossary, or help from other students. Therefore, Brady was unconvinced that the strategy of searching the text to clarify the

meanings of words was of any help in the typical social studies texts that he was using. On the other hand, Brady did note that one positive effect of encouraging students to ask for clarification of a word was that it legitimized students' efforts to seek help from and offer assistance to each other. Mutual help might be part of what students like about reciprocal teaching.

Brady noted that students also had difficulty using the prediction strategy. He wrote that "it proved virtually impossible to accurately predict the content of the next paragraph" (pp. 104–105), partly because the social studies text he used lacked coherence, and partly because of the chronological nature of the text. Unfortunately, no other investigator discussed problems that occurred in actually attempting to apply the four strategies.

In summary, we are uncertain which are the more effective strategies to use with reciprocal teaching, and how many strategies should be taught. The two strongest candidates are question generation and summarization. As Palincsar and Brown (1984) have written, asking questions and summarizing serve a comprehension-fostering function in that they require students to search text and perform deeper processing. These activities also have a comprehension-monitoring function in that difficulty in performing either task signals the learner that there are comprehension difficulties. More information is needed about changes in student processing that occur as a result of strategy instruction.

The Teaching of Cognitive Strategies

The studies in reciprocal teaching also allow us to discuss a number of practical issues regarding the specific teaching of cognitive strategies. These topics include three general approaches for teaching cognitive strategies, procedural prompts that were used to help students learn the strategies, a listing of specific instructional procedures that were used when teaching cognitive strategies, suggestions for improving the quality of the dialogue in reciprocal teaching, and a discussion about teaching below-average students.

Three Instructional Approaches for Teaching Cognitive Strategies

Investigators have developed a number of overlapping approaches for teaching students cognitive strategies that may help improve their comprehension. Two approaches appear in this review. In the first approach, all instruction in cognitive strategies takes place *during* the reciprocal teaching dialogues, in response to specific student problems (RTO). In the second, the strategies are taught explicitly *before* beginning the dialogues (ET-RT). In the reciprocal teaching studies, both the RTO and ET-RT approaches were usually successful when experimenter-developed comprehension tests were used. Neither approach was very effective when standardized tests were used.

There is a third, overlapping instructional approach to teaching cognitive strategies that might be termed *general cognitive strategy instruction*. In this approach, the cognitive strategies are first taught explicitly, as in the ET-RT approach. Then the teacher guides the students as they practice applying the strategy. The teacher gradually withdraws this support as the students become more competent. Models, guided practice, checklists, and thinking aloud are used in these studies to help or scaffold the students, but there is no reciprocal teaching.

This third, general approach was used in the studies cited at the beginning of this article—studies in reading comprehension (Alvermann, 1981; Paris et al., 1984; Raphael & Pearson, 1985), mathematics problem solving (Schoenfeld, 1985), physics problem solving (Larkin & Reif, 1976), and writing (Englert & Raphael, 1989; Scardamalia & Bereiter, 1985). The description of the following five studies in reading provides more information on the instructional procedures in this general cognitive strategy approach.

Davey and McBride (1986) used prompts, models, guided practice, and a self-evaluation checklist to help students learn to generate questions. Students then worked alone generating questions on passages and applying the self-evaluation checklist. In a study by Cohen (1983), students were provided with models and guided practice, and then worked alone to complete “question generating booklets.” Blaha (1979) provided students with three questions that were intended to help them summarize a passage. The students were first taught in a group to generate and answer these questions, and then they practiced generating and answering questions individually. Finally, the students came together in a group to compare answers and clear up difficulties.

Bereiter and Bird (1985) taught cognitive strategies through a series of steps: modeling, thinking aloud, labeling the strategy, and then reducing the support given to the students. First, the instructor performed all the steps for the students: she read passages aloud, and then thought aloud as she applied strategies to resolve comprehension difficulties and labeled the strategies. Then the instructor withdrew her support by reading new passages aloud and applying the strategies while requiring the students to label the strategies being used. In the next step, the students worked alone, read passages, applied the strategies and orally marked their use of each strategy. In other words, first the teacher modeled and identified a strategy, then the teacher modeled the strategy and the students identified it, and finally, students both used and labeled the strategy.

A similar instructional procedure of modeling, labeling, and reducing the role of the teacher while increasing the role of the students was used by Raphael and Pearson (1985) as they taught students to identify different types of questions and the different cognitive processes each type required. The investigators first modeled the questions and identified them by type. Next, the investigators modeled the questions and students were required to identify them by type. Finally, the students generated the questions and identified their type.

In all five studies, students learned cognitive strategies through explicit instruction, but without the reciprocal teaching framework. In all five, students in the experimental group had significantly higher posttest scores than students in the control group. In four of the studies, the dependent measures were experimenter-developed comprehension tests (Blaha, 1979; Cohen, 1983; Davey & McBride, 1986; Raphael & Pearson, 1985); in the fifth study, the dependent measure was a standardized test (Bereiter & Bird, 1985). These studies are fairly typical of studies that use a general cognitive strategy approach.

There are many common elements in the above three approaches (i.e., reciprocal teaching only, explicit teaching before reciprocal teaching, and general cognitive strategy instruction). All three approaches use modeling, provide scaffolds for the students, guide and support student practice, require students to respond, provide feedback for their responses, and diminish the teacher support and peer

support as students become more competent. The approaches differ in *when* the cognitive strategies are taught, how much explicit instruction is provided, and amount of peer support. The general approach and the ET-RT approach provide explicit instruction at the start, whereas, in the RTO approach, instruction occurs only when there is a need.

At present, all three approaches have yielded significant results, particularly when experimenter-developed comprehension tests have been used. It may be useful to study which specific instructional elements in these approaches are most effective, and develop theoretical explanations to account for their effects.

Procedural Prompts for Teaching Cognitive Strategies

When investigators used the ET-RT approach, they usually provided students with scaffolds in the form of specific procedural prompts or facilitators (Scardamalia & Bereiter, 1985) to help them learn and apply strategies. For example, in at least five of the reciprocal teaching studies students were provided with *signal words* such as “when,” “why,” and “how,” and were taught to use these words as prompts to help generate questions. At least three different procedural prompts were used in these studies to help students learn the strategy of question generation, and these are listed in Table 15. One of these prompts, question type, was used to show students that there were three types of questions—those with answers that lay in a single sentence, those with answers that required combining information, and those with answers that lay outside the text (Dermody, 1988). Students used this concept to help generate questions. Students were also taught how to use main ideas as an aid in generating questions (Lonberger, 1988). Each procedural prompt was used in at least one study that obtained significant results. Unfortunately, there were too few studies using each type of prompt to enable us to discuss whether any one prompt was more successful than the others.

Two different procedural prompts were used to help students apply the cognitive strategy of summarization. In one approach, used by Palincsar (1987) and Brady (1990), students were taught to apply a set of procedures that were derived from

TABLE 15

Prompts for questioning and summarizing taught in different reciprocal teaching studies

Question Generation	
Signal words	Brady, 1990; Lysynchuk et al., 1990; Palincsar, 1987; Palincsar & Brown, 1984; Taylor & Frye, 1992; Williamson, 1989
Question types	Dermody, 1988
First locate main idea	Lonberger, 1988; Rich, 1989
Summarization	
Kintsch-van Dijk procedures	Palincsar, 1987; Brady, 1990
Legs-and-table procedures	Taylor & Frye, 1992; Dermody, 1988; Lonberger, 1988
“Think and construct”	Lysynchuk et al., 1990
None found	Palincsar & Brown, 1984; Levin, 1989; Williamson, 1989

the procedures used by experts to summarize text. These procedures were first identified by Kintsch & van Dijk (1978).

Step 1. Identify the topic sentence.

Step 2. If a topic sentence is not given, then identify the topic and the most important information about that topic.

Rule 1. Leave out unimportant information.

Rule 2. Give steps or list a title.

Rule 3. Cross out information that is redundant/repeated.

Brady noted that during the pilot testing for his study students often made general statements instead of summaries; for instance, a student might say, "This paragraph was about horses," or "This paragraph is about herds." To avoid that problem, Brady told students to use the following sentence starter: "This paragraph tells us that. . ."

Other investigators (Dermody, 1988; Lonberger, 1988; Taylor & Frye, 1992) taught students a third procedural prompt for summarizing. Students were taught to first list the main details in a paragraph, and then use these details as a prompt to generate a summary.

In both of the above examples, the strategy of summarization was not taught holistically. Rather, each component was taught separately and was followed by student practice. Then the teacher helped the students as they consolidated these components, and, finally, the teacher withdrew this support as the students became independent. In summary, the strategies for summarization were taught as a series of cognitive subskills.

There were four other studies that did not describe any prompts for teaching summarization (Levin, 1989; Lysynchuk et al., 1990; Palincsar & Brown, 1984; Williamson, 1989). Two of these studies obtained significant results (Levin, 1989; Palincsar & Brown, 1984), one obtained mixed results (Lysynchuk et al., 1990), and one obtained nonsignificant results (Williamson, 1989).

We looked at the facilitators or prompts that were used to teach the two major cognitive strategies—question generation and summarization—in hopes of identifying the more effective prompts. We were somewhat surprised to find that in the context of these studies, all of the prompts used to teach question generation and summarization were equally successful. Indeed, 9 of the 10 studies that taught cognitive strategies using procedural prompts were successful in at least one of the two types of outcome measures. It will be useful to conduct future experimental studies that contrast the use of different facilitators for teaching the same cognitive strategy.

In sum, a variety of scaffolds in the form of procedural prompts successfully supported students as they learned the cognitive strategies. It is possible that different procedural prompts can be equally successful, and there is a need for further study on the characteristics of successful procedural prompts. There is also a need for study to determine when, for which cognitive strategies, and for what type of student, these procedural prompts are needed.

Other Instructional Procedures That May Have Aided the Teaching of Cognitive Strategies

In addition to procedural prompts, a number of other interesting instructional procedures were used in some of the reciprocal teaching studies. These are described below.

1. Students were provided with cue cards that contained procedural prompts to which they could refer during practice (Brady, 1990).
2. The process of using the procedural prompts to develop questions and summaries was modeled by the teacher. Students were also provided with models of good questions and summaries.
3. The difficulty of the instructional task was regulated for the students by selecting practice materials that were below the grade level of the students (Levin, 1989; Lonberger, 1988; Rich, 1989).
4. The difficulty of the instructional task was regulated by starting with simpler tasks, such as generating questions about a single sentence, and then moving to questions about entire paragraphs (Brady, 1990; Palincsar, 1987). The difficulty of the task was also regulated by having the teacher model the more difficult part of the task as the students carried out those parts of the task they were capable of completing.
5. Students were provided with self-checking or fix-up procedures to which they could refer after completing the task (Lonberger, 1988).
6. Student responsibility was gradually increased, although the teacher was available to support students as they performed certain aspects of the task (Lonberger, 1988).

Quality of the Dialogue in Reciprocal Teaching

Dialogue plays a critical role in providing scaffolded instruction (Palincsar, 1986). However, most of the investigators did not evaluate the quality of the dialogue in their studies, and, as we noted earlier, Palincsar and Brown, in their original 1984 study, did not provide a checklist of criteria that could be applied to evaluate reciprocal teaching dialogues.

A few years later, Palincsar and her associates (Palincsar, 1986; Palincsar et al., 1989) attempted to describe the important instructional features of dialogues by comparing the dialogues of several primary teachers (Palincsar, 1986; Palincsar et al., 1989). They found that the less successful teachers often merely recited information about the strategy being taught during the dialogues. Students' involvement occurred mostly at the word level; they were asked only to provide labels for strategies or to complete sentences begun by the teacher. In contrast, the more successful teachers engaged students at the "idea" level and focused the dialogue on using the strategies when discussing a passage. They encouraged students to engage in the dialogue by rephrasing questions when students were unable to respond, and by using and elaborating upon other students' responses. They provided support by explicitly modeling strategies ("If I were asking a question, I . . ."; "If I were going to summarize, I . . ."), and helped students label the language or features of the interaction.

The following features, which emerged in Palincsar's (1986) analysis as distinguishing certain dialogue, might be used to evaluate implementation of reciprocal teaching in future studies:

- (a) teacher support of the students' contribution to the dialogue at the idea level as opposed to the word level;
- (b) deft use of student ideas and linking of those ideas to new knowledge;
- (c) dialogues with focus and direction;

- (d) the purpose of the instruction was made explicit to the student and seemed explicit to the teacher; and
- (e) evaluative statements were made that changed the complexion of a student response from negative to constructive. (p. 96)

Teaching Below-Average Students

In their original study, Palincsar and Brown (1984) selected only those students who were near grade level in decoding but below grade level in comprehension. This selection criterion implied that reciprocal teaching was not likely to be as successful with students who are poor in decoding. Six reciprocal teaching studies used below-average students and did not screen for decoding ability; we assume that within this group there were students who had poor decoding skills. Two of these six studies used experimenter-developed comprehension tests, and both obtained significant results. Only one of the five studies that used standardized tests obtained significant results. This pattern with below-average students, then, is quite similar to the results obtained in studies that used good-poor or all students. However, we cannot say how well reciprocal teaching worked with students who were below-average in decoding, because the investigators did not present separate results for those students.

We looked at all six studies that taught below-average students to see if we could discern common elements across the three studies that obtained significant results. We looked at the reading levels of the materials and the types of instruction but, because of the small number of studies, we were unable to find any pattern.

Future Practice and Research

This final section on future research begins with a discussion of procedures for conducting, reporting, and reviewing intervention studies. It closes with a discussion of five possible research questions on reciprocal teaching and the teaching of cognitive strategies.

Suggestions for Conducting and Reporting Future Studies

We would like to make a few recommendations regarding the conducting, reporting, and reviewing of future studies which might help to clarify some of the questions we have raised. The work of Palincsar and Brown (1984), and Brady (1990), was particularly useful for writing this section.

Assessing student learning during the study. A number of investigators administered various assessments both during and after the studies. These assessments included (a) monitoring student comprehension during the study (Palincsar and Brown, 1984; Brady, 1990), (b) evaluating student learning of the strategies after the intervention had been completed (Palincsar and Brown, 1984), (c) providing data on student scores on each type of question (Brady, 1990), and (d) conducting follow-up testing 60 days after the posttest (Brady, 1990). We hope that these assessments will be included in future studies in order to provide more complete results on the effects of the intervention.

Assessing implementation. One of the shortcomings of these and other intervention studies is that little attention has been paid to how well the intervention was taught. In reciprocal teaching studies, for example, there was little assessment

of the quality of the dialogues. No reciprocal teaching study provided a checklist that could be used to evaluate the quality of the dialogues using criteria that were specific to reciprocal teaching. Few studies provided samples of the actual dialogues. Aside from the original Palincsar and Brown (1984) study, there was little assessment of the quality of the student questions and summaries during the dialogues. Those who wish to implement reciprocal teaching in their classrooms, or who wish to conduct future studies, have no source for guidance in practice and for assessment of implementation. We hope that the earlier sections on monitoring student learning of the strategies and our list of the implementation items in Table 12 can contribute to the development of such a source, but we note there is a need for increased work in developing assessment procedures.

Analyzing results by ability level. We were surprised that in studies involving all the students in classes, investigators performed little analysis of results by entering ability level. Such analysis would help researchers and reviewers determine if all students benefit equally from reciprocal teaching.

Discussing problems that occur during the dialogues. We were impressed with Brady's discussion of the problems that occurred in dialogues when different strategies were tried. Future development of similar studies might be improved if researchers would identify and discuss problems that arise during instruction and dialogues.

Suggestions for Reviewing Intervention Studies

Assessing the quality of the instruction. When conducting reviews such as this one, on instructional interventions, it is important to evaluate not only the design of each study, but also the quality of the instruction actually given. However, we found it difficult to make this evaluation. Half the studies, those in which experimenters gave instruction, did not evaluate or discuss the quality of that instruction. Only two investigators provided us with sample transcripts of lessons, and few provided detailed description of the instruction that we could use to make an evaluation. Indeed, evaluating the quality of instruction does not appear to be common practice in the reporting or reviewing of intervention studies. This lack of evaluation of the quality of the instruction even appears in a methodological analysis of the *quality* of experimental studies of comprehension strategy instruction (Lysynchuk, Pressley, d'Ailly, Smith, & Cake, 1989). In this evaluation, the authors developed 24 "criteria for internal validity" which they used to evaluate 37 studies. These criteria included instructional variables, such as the amount of time students spent on tasks, but none focused on the quality of instruction in an intervention study.

We believe that, in the future, researchers should develop initial criteria for evaluating the quality of instruction and should address the quality of instruction in their results and discussion sections. At the very least, editors, reviewers, and advisers should insist that the report of a study contain full descriptions of instructional procedures, with particular emphasis on the amount and type of support—guided practice and scaffolds—that was provided for students. This information would allow readers and reviewers to address what effects, if any, the quality of instruction had on the results.

Analyzing results by type of test. One finding in this review was that investigators found much higher median effect sizes with experimenter-developed comprehension tests than with standardized tests. We looked at other cognitive strategy studies, those that taught the strategy of question generation but did not use the reciprocal teaching approach, and found the same result. Thus, we would recommend that in future reviews researchers first determine whether different types of tests, even different experimenter-developed comprehension tests, are yielding different results.

Future Research Questions

Why Are Experimenter-Developed Comprehension Tests and Standardized Tests Producing Different Results?

Two types of outcome measures were used in these studies: standardized reading tests and experimenter-developed tests. Only 2 of the 13 studies that used standardized reading tests to assess comprehension had significant results, whereas 8 of the 10 studies that used experimenter-developed comprehension tests showed significant results. Further, in 4 of the 5 studies that used both types of test, the results were significant when experimenter-developed tests were used and nonsignificant when standardized tests were used. It appears, then, that significant results were obtained much more frequently when experimenter-developed tests were used.

Earlier, we compared the Gates-MacGinitie (MacGinitie, 1978) reading test with the passages and questions developed by Palincsar and found a number of differences that we believe made the Gates-MacGinitie test more difficult. We recommend future study of (a) the differences between these two types of test and (b) the implications for instruction.

How Do the Internal Cognitive Processes Used by Students Change as a Result of Instruction?

Although we can note the effect of teaching cognitive strategies on student comprehension scores, we do not know what internal cognitive processes changed as a result of instruction and practice. This issue seems worthy of future research, perhaps using procedures similar to the “think aloud” method that Bereiter and Bird (1985) used to study the internal processes of expert readers.

What Are the Critical Instructional Elements for Teaching Cognitive Strategies?

Three general approaches toward teaching cognitive strategies have been identified: instruction only during dialogues (RTO), explicit teaching before dialogues began (ET-RT), and a general approach involving presentation, guided practice, and independent practice but without dialogues. We do not endorse studies that purport simply to compare these approaches because there are too many variables within each approach. Rather, we recommend the study of specific instructional elements in these approaches to determine which are most effective, and the development of theoretical explanations to account for these effects. Some instructional elements which appear useful across approaches and which might be studied are listed above in the section on instructional procedures.

Which Procedural Prompts Are Most Useful for Teaching Cognitive Strategies?

We looked at the prompts that were used to teach the two major cognitive strategies—question generation and summarization—in hopes of identifying the more effective prompts. The different prompts and the studies in which they were used are presented in Table 15. We found that when experimenter-developed comprehension tests were used, all of the prompts used to teach question generation and summarization were equally successful. The question, then, of which prompts are more useful for teaching cognitive strategies awaits future experimental studies that contrast the use of different prompts for teaching the same cognitive strategy.

How Many and Which Strategies Are Most Productive?

In these 16 reciprocal teaching studies, investigators achieved significant gains by teaching from 2 to 10 cognitive strategies. A number of studies outside the reciprocal teaching tradition that taught only single strategies also obtained significant results. Future investigations might focus on the effects of teaching individual strategies and combinations of strategies.

Coda

There are at least five excellent instructional ideas embedded in reciprocal teaching: (a) the focus on helping students acquire comprehension-fostering strategies instead of simply asking them comprehension questions, (b) the provision of four specific comprehension-fostering strategies instead of the tens of “reading skills” that have appeared in reading workbooks, (c) the provision for practicing the strategies while reading actual text, (d) the popularization of procedures for scaffolding or supporting students as they develop their strategies, and (e) the popularization of the idea of students providing support for each other within reading groups. These ideas, particularly (a), (b), and (c), have existed before, but the reciprocal teaching method packages and presents them in a manner that has gained educational acceptance.

We suggest that the main weakness in the practice of reciprocal teaching is that not enough has been written on implementation. There is no checklist of criteria for assessing the quality of reciprocal teaching instruction, and we found it difficult to evaluate the quality of the reciprocal teaching in the studies we reviewed.

We view reciprocal teaching as an example of cognitive strategy instruction, and based on the generally favorable record of that research, we recommend that such instruction become part of ongoing practice. We also note that there are a wide variety of instructional issues within cognitive strategy instruction, and we hope there will be future research on these topics.

APPENDIX A

Studies that used good-poor/all students

Study	Length	Level	Type of student	Dependent measure	Sign.	Effect Size
Reciprocal Teaching Only						
Jones, 1987	20 sessions	3rd	all	Informal reading inventory	ns	.34 ^a
Palincsar & Brown, 1984	20 sessions	7th	poor comp./ good decoders	comp. quest. (open-ended)	S	1.0 + (est.)
Rush & Milburn, 1988	9 sessions	vocat. college	all	Cooperative English test	ns	no s.d.; 4 treatments
Explicit Teaching Before Reciprocal Teaching						
<i>Signal Words Facilitator</i>						
Lysynchuk et al., 1990	13 sessions	4th and 7th	poor comp./ good decoders	Gates-MacGinitie Metropolitan comp. test	ns ns/S	.55 .68
Palincsar, 1987	25 sessions	middle school	poor comp./ good decoders	summarizing (levels) comp. ques. (open-ended)	S S	.68 (est.) 1.08 (est.)
Taylor & Frye, 1992	11 sessions	5th and 6th	average and above	Gates-MacGinitie summarizing (levels)	ns S	.07 .85
Williamson, 1989	50 sessions	3rd	all	IL State Assess.	ns	.32

APPENDIX A—continued

Study	Length	Level	Type of student	Dependent measure	Sign.	Effect Size
<i>Question Type Facilitator</i>						
Dermody, 1988	24 sessions	4th	all	Stanford Diag. Reading Test summarizing (total prop.)	ns	-.32
					S	3.37
<i>Main Idea Facilitator</i>						
Lonberger, 1988	20 sessions	4th and 6th	all	CR test (exper. developed)	S	1.24
<i>Graphic Organizer</i>						
Shortland- Jones, 1986	20 sessions	Grades 1-6	mixed (summer reading program)	Stanford Diag. Reading Test summarizing	S	.77
					ns	-.02

^aAverage of two experimental groups

APPENDIX B

Studies that used below-average students

Study	Length	Level	Type of student	Dependent measure	Sign.	Effect size
Reciprocal Teaching Only						
Fischer Galbert, 1989	100 sessions ^a	3rd, 4th, and 5th	Chapter I	Iowa TBS	ns ^b	-.21 (est.)
Labercane & Battle, 1987	28 sessions	5th	learning disabled	Gates-MacGinitie	ns	.36 (est.)
Padron, 1985	8 sessions	8-12	Hispanic	Stanford DRT (comp. subtest)	ns	-.55
Rich, 1989	6 sessions	year olds adult	bilingual poor readers	Mult. choice comprehension Recall (idea units)	S	1.74 ^c
					S	1.10 ^d
Explicit Teaching Before Reciprocal Teaching						
<i>Signal Words Facilitator</i>						
Brady, 1990	25 sessions	7th	below-average (native Alaskan)	Comp. test (open-ended)	S	.87
				Gates-MacGinitie	ns	.36
<i>Informed Strategies for Learning</i>						
Levin, 1989	50 sessions	intermed.	learning disabled	Stanford Achiev. comp. subtest	S	no adj. means

^aNumber represents the total number of Chapter I classes

^bMay favor control

^cAverage of three experimental groups

^dAverage of three experimental groups

APPENDIX C

The following is a script illustrating the explicit teaching of questioning prior to the reciprocal teaching dialogue. This script was developed by Annemarie Palincsar.

“Questions play an important part in our lives. Much of our class discussion is focused on answering questions. What are some other occasions or situations when questions are important?”

Elicit responses from the students that might include the following:

- Reading assignments usually require us to answer questions at the end of a story.
- Tests usually require us to answer a series of questions.
- When we need more information about something, we need to ask questions.

“Let’s practice asking some questions for situations when we might need more information. For example, suppose you want to see the movie, ‘The Empire Strikes Back,’ however, you don’t know when the movie begins. You might call the theater and ask ‘*What time does the movie, “The Empire Strikes Back” begin?*’

“One of the activities we will be doing for the next couple of weeks is learning to ask good *questions* about material we have read. We will focus on asking questions about important information rather than about unimportant, trivial, or detailed information.

“There are several reasons why we will learn to ask questions while reading: (list italicized phrases on the chalkboard)

- (a) It is a way in which we can *test ourselves* to make sure we understand what we have read.
- (b) It is a good way to *focus on important information* in a passage.
- (c) With a little practice it is possible that we can become skilled enough at questioning that we can *predict the kinds of questions* we might be asked on a test. This would be very useful while studying.

“Let’s begin by talking about the words that are used to ask questions. What are some of the words that we use to begin sentences that are questioning sentences?” (List responses on the chalkboard.)

Who	What
When	Where
Why	How

“Let’s practice by asking questions about the following sentences. At first you will be given the question word; however, later you will be asked to think of your question words. Look at the first sentence on your papers. The falcon is a female hunting bird.¹ Ask a question word about the information in this sentence that begins with the word ‘what.’ ”

1. The falcon is a female hunting bird.

WHAT is a falcon? or **What is the name of a female hunting bird?** or **What does a falcon do?** (Accept any appropriate responses.)

“Ask a question word about the information in Sentence 2. Begin your question with the word ‘who.’ ”

2. In medieval times, in Europe, only members of a royal family could own falcons.

WHO could own falcons in medieval times?

“Ask a question about the third sentence that begins with the word ‘why.’ ”

3. The falcon bathes in shallow streams to control bird lice that live in her feathers.

WHY does the falcon need to bathe?

4. A falcon prefers to hunt for its prey in open areas.

WHERE do falcons hunt?

5. In the 1950s the falcon populations in North America and Central Europe dropped suddenly.

WHEN was there a decline in the falcon population?

6. The falcon hunts by swooping down on her prey and grabbing it with her sharp talons.

HOW does the falcon catch her prey?

“For further practice, make up questions for each of the following sentences (7–11). This time, however, no question words are given.” You may continue to complete these items as a whole-class discussion or give the students the opportunity to complete them independently by writing down their responses. Discuss their questions and accept any question that captures the main idea of the sentence and is posed clearly.

7. Although animals don’t have language as we do, they do communicate with each other by signals of some kind.
8. Scientists study animal communication through experiments and observations.
9. Because snakes are totally deaf, it is the movement of the snake charmer that charms the snake, not the music the snake charmer plays.
10. Some ants give off a special alarm odor that warns nearby ants of danger.
11. The sounds made by bats, moths, and whales are too high for humans to hear.

“Now that you are successful making up questions, we will discuss selecting the most important information in the paragraph about which to ask a question. Look at number 12 on your papers. Number 12 is a short paragraph. After the paragraph there are three questions. One of these questions is better than the other two because it is about the most important information in the paragraph. Let’s first read the paragraph.”

12. Deaths from snakebite have been cut down in recent years by the use of antivenins—medicines that work against the snake poisons. There are now few deaths from snakebite in the United States and Canada.

“Let’s read the three questions that follow this paragraph and try to decide which question asks about the most important information.” (Read through all three choices.)

_____ a. Why do snakes bite people?

“This is not a good main idea question. Can anyone tell us why?” (This is not a good main idea question because the question is not answered in the paragraph. This question would make a good prediction, however, since it shows that you are thinking about what kinds of information might come next in the story.)

_____ b. In what countries do few people die from snakebite?

“This is what we could call a detail question. While the answer is in the paragraph, it is not about the most important information in the paragraph.”

_____ c. Why do fewer people die from snakebite these days?

“This question is the best because to answer it you must discuss the antivenin or medicine which is the main topic of the paragraph.”

“Let’s try another example. Read Paragraph 13 and the three questions that follow it. Put a check mark next to the best question. Remember that the best question should be clear and should be about the most important information in the paragraph. Be ready to discuss your choice.” Once students have selected the best response, discuss each question and why it is/is not the best choice.

13. Contrary to what some people believe, snakes do not sting with their tongues. Their tongues are used to sharpen their sense of smell. The snake picks up tiny particles of matter in the air with his tongue and puts them in two tiny holes at the bottom of his nostrils so that he can smell better.

_____ a. How many holes does a snake have at the bottom of his nostrils? (detail question)

- _____ b. What does a snake use his tongue for? (main idea question since paragraph discusses how a snake uses his tongue for smelling rather than for stinging)
- _____ c. Why do people use the expression, "He speaks with forked tongue"? (question not answered in the text)

Instruct students to select the best main idea questions for Paragraphs 14 through 16. Tell them to be ready to defend their answer by telling the reasons for their choice. Discuss student responses.

14. The smallest snake is just about the size of a worm. The largest snake has been known to reach thirty feet in length, which is almost as long as two station wagons. There are many varieties of snakes and they come in many different lengths.
- _____ a. How long do snakes get? (all sentences discuss length of snakes)
 - _____ b. How many station wagons could fit into 30 feet? (not even mentioned in paragraph)
 - _____ c. Where would you find the longest snake? (not discussed in paragraph; a question like this would make a good prediction for the next portion of text)
15. Snakes are very flexible because their body is like a rubber hose with many bones. In fact, a snake's backbone can have as many as 300 vertebrae, almost ten times as many as a human's. Because of all these bones, a snake can twist its body in almost any direction.
- _____ a. Why can a snake move its body in so many ways? (entire paragraph is about flexibility and movement of snakes)
 - _____ b. Do snakes ever need backrubs the way people do? (not mentioned)
 - _____ c. How many vertebrae do snakes have? (detail question)
16. While very small snakes eat very small insects or worms, large snakes can eat small deer, leopards, and goats. All snakes, regardless of size, eat living animals or animal eggs. In fact, some snakes swallow each other.
- _____ a. What snake eats its neighbor snake? (not mentioned in paragraph)
 - _____ b. What do snakes eat? (good main idea question since entire paragraph describes what snakes eat)
 - _____ c. How is the diet of the small snake different from the diet of the snake? (a good question; however, this question is only about one sentence rather than the entire paragraph)

Instruct students to write main idea questions for Paragraphs 17 through 20. Discuss responses.

17. Camels have been helpful to people who live in deserts for thousands of years. They have carried people as well as their goods on their strangely shaped backs. They are able to cross deserts and mountains on trips that may take two months. (*Why or how are camels helpful to people?*)
18. Scientists have studied the camel carefully to determine how it can live where other animals would die. They have found that the camel's body is especially well designed for its life in the hot, dry, sandy parts of the world. There are many characteristics of the camel that are useful to it including its feet, legs, eyelashes, and nostrils. (*Why is the camel able to live in the desert?*)
19. There have been many prominent women in America's history who have done much good for mankind. One of these women was Alice Hamilton. Dr. Hamilton was very concerned about the health of industrial workers. Through her research and leadership she was responsible for many changes that improved working conditions for laborers. (*What did Dr. Alice Hamilton do to help people? or What was important about Dr. Alice Hamilton?*)

20. Scientists have been asking themselves what energy is for hundreds of years, but no one has come up with a simple answer. About the only definition of energy that scientists can agree on is that energy is that something which enables people, machines, and objects to do work.
(What is energy?)

Source: Palincsar, 1987

References

- Alvermann, D. E. (1981). The compensatory effect of graphic organizers on descriptive text. *Journal of Educational Research*, 75, 44–48.
- Anderson, L., Evertson, C., & Brophy, J. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal*, 79, 193–223.
- Andre, M. D. A., & Anderson, T. H. (1979). The development and evaluation of a self-questioning study technique. *Reading Research Quarterly*, 14, 605–623.
- Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction*, 2, 131–156.
- Bird, M. (1980). *Reading comprehension strategies: A direct teaching approach*. Unpublished doctoral dissertation, University of Toronto, Toronto.
- Blaha, B. A. (1979). *The effects of answering self-generated questions on reading*. Unpublished doctoral dissertation, Boston University.
- Brady, P. L. (1990). *Improving the reading comprehension of middle school students through reciprocal teaching and semantic mapping strategies*. Unpublished doctoral dissertation, University of Oregon–Eugene.
- Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher*, 10(2), 14–21.
- Brown, A. L., & Palincsar, A. S. (1985). *Reciprocal teaching of comprehension strategies: A natural history of one program for enhancing learning*. (Tech. Rep. No. 334). Urbana-Champaign: University of Illinois, Center for the Study of Reading.
- Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393–451). Hillsdale, NJ: Erlbaum.
- Cohen, R. (1983). Self-generated questions as an aid to reading comprehension. *The Reading Teacher*, 36, 770–775.
- Dansereau, D. F. (1980, October). *Learning strategy research*. Paper presented at NIE-LRDC Conference on Thinking and Learning Skills, University of Pittsburgh.
- Dansereau, D. F., Collins, K. W., McDonald, B. A., Halley, C. D., Garland, J. E., Diekkoff, G., & Evans, S. H. (1979). Development and evaluation of an effective learning strategy program. *Journal of Educational Psychology*, 71(1), 64–73.
- Davey, B. (1983). Think aloud: Modeling the cognitive process of reading comprehension. *Journal of Reading*, 27, 44–47.
- Davey, B., & McBride, S. (1986). Effects of question-generation training 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.
- Duffy, G., Lanier, J. E., & Roehler, L. R. (1980). *On the need to consider instructional practice when looking for instructional implications*. Paper presented at the Conference on Reading Expository Materials, Wisconsin Research and Development Center, University of Wisconsin–Madison.
- 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. I), (pp. 105–151). Newark, NJ: JAI Press.
- Fischer Galbert, J. L. (1989). *An experimental study of reciprocal teaching of expository text with third, fourth, and fifth grade students enrolled in Chapter 1 Reading*. Unpublished doctoral dissertation, Ball State University, Muncie, IN.
- Gilroy, A., & Moore, D. (1988). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities with ten primary school girls. *Educational Psychology*, 8, 41–49.
- Good, T., & Grouws, D. (1979). The Missouri teacher effectiveness program. *Journal of Educational Psychology*, 71, 355–362.
- Greenfield, P. M. (1984). A theory of the teacher in the learning activities of everyday life. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 117–138). Cambridge, MA: Harvard University Press.
- Helfeldt, J. P., & Lalik, R. (1976). Reciprocal student-teacher questioning. *Reading Teacher*, 33, 283–287.
- Hunter, M. (1982). *Mastery Teaching*. El Segundo, CA: TIP Publications.
- Jones, M. P. (1987). *Effects of reciprocal teaching method on third graders' decoding and comprehension abilities*. Unpublished doctoral dissertation, Texas A&M University.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- 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.
- Lave, J. (1977). Tailor-made experiences in evaluating the intellectual consequences of apprenticeship training. *Quarterly Newsletter of Institute for Comparative Human Development*, 1, 1–3.
- Levin, M. C. (1989). *An experimental investigation of reciprocal teaching and informed strategies for learning taught to learning-disabled intermediate school learners*. Unpublished doctoral dissertation, Teachers College, Columbia University.
- 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., d'Ailly, H., Smith, M., & Cake, H. (1989). A methodological analysis of experimental studies of comprehension strategy instruction. *Reading Research Quarterly*, 24, 458–470.
- Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *Elementary School Journal*, 90, 471–484.
- MacGinitie, W. H. (1978). *Gates-MacGinitie reading tests: Level D*. (2nd ed.) Boston: Houghton Mifflin.
- Manzo, A. V. (1969). *Improving reading comprehension through reciprocal questioning*. Unpublished doctoral dissertation, Syracuse University.
- Meichenbaum, D. (1977). *Cognitive behavior modification: An integrative approach*. New York: Plenum.
- Mosenthal, J. H. (1987). Learning from discussion: Requirements and constraints on classroom instruction in reading comprehension strategies. In J. Readence & S.

- Baldwin (Eds.), *Research in literacy: Merging perspectives* (pp. 169–176). Rochester, NY: National Reading Conference.
- 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.
- Padron, Y. N. (1985). *Utilizing cognitive reading strategies to improve English reading comprehension of Spanish-speaking bilingual students*. Unpublished doctoral dissertation, University of Houston.
- Palincsar, A. S. (1982). *Improving the reading comprehension of junior high students through the reciprocal teaching of comprehension-monitoring strategies*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Palincsar, A. S. (1986). The role of dialogue in providing scaffolded instruction. *Educational Psychologist*, 21, 73–98.
- 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. (1983). *Reciprocal teaching of comprehension-monitoring activities*. (Tech. Rep. No. 269). Urbana-Champaign: University of Illinois, Center for the Study of Reading.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 2, 117–175.
- Palincsar, A. S., Brown, A. L., & Martin, S. M. (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist*, 22, 231–253.
- Palincsar, A. S., David, Y., Winn, J., Stevens, D., & Brown, A. L. (1990, April). Examining the differential effects of teacher- versus student-controlled activity in comprehension instruction. Paper presented to the annual meeting of the American Educational Research Association, Boston.
- Palincsar, A. S., Stevens, D. D., & Gavelek, J. R. (1989). Collaborating with teachers in the interest of student collaboration. *International Journal of Educational Research*, 13, 41–53.
- 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.
- Raphael, T. E., & Pearson, P. D. (1985). Increasing student awareness of sources of information for answering questions. *American Educational Research Journal*, 22, 217–237.
- Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academic Press.
- Rich, R. Z. (1989). *The effects of training adult poor readers to use text comprehension strategies*. Unpublished doctoral dissertation, Teachers College, Columbia University.
- Risko, V. J., & Feldman, N. (1986). Teaching young remedial readers to generate questions as they read. *Reading Horizons*, 27, 54–64.
- Rogoff, B., & Gardner, W. P. (1984). Adult guidance of cognitive development. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 95–116). Cambridge, MA: Harvard University Press.
- Rosenshine, B., & Berliner, D. (1978). Academic Engaged Time. *British Journal of Teacher Education*, 4, 3–16.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.) *Handbook of Research on Teaching* (3rd ed.) (pp. 376–392). New York: Macmillan.

- Rush, R. T., & Milburn, J. L. (1988, November). *The effects of reciprocal teaching on self-regulation of reading comprehension in a postsecondary technical school program*. Paper presented at the annual meeting of the National Reading Conference, Tucson, AZ.
- 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: Research and open questions*. Hillsdale, NJ: Erlbaum.
- Schoenfeld, A. H. (1985). *Mathematical Problem Solving*. New York: Academic Press.
- Shortland-Jones, B. (1986). *The development and testing of an instructional strategy for improving reading comprehension based on schema and metacognitive theories*. Unpublished doctoral dissertation, University of Oregon.
- Slavin, R. E. (1987). Mastery learning reconsidered. *Review of Educational Research*, 57, 175–215.
- Taylor, B. M., & Frye, B. J. (1992). Comprehension strategy instruction in the intermediate grades. *Reading Research and Instruction*, 92, 39–48.
- Vygotsky, L. S. (1962). *Thought and Language* (E. Hanfmnn and G. Vahar, Eds. and Trans.) Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Schribner, & E. Souberman, Eds. and Trans.). Cambridge, MA: Harvard University Press.
- Wertsch, J. V., & Stone, C. A. (1979, February). *A social interaction analysis of learning disabilities remediation*. Paper presented at the International Conference of the Association for Children with Learning Disabilities, San Francisco.
- 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, Y. L., & Jones, W. (1982). Increasing metacomprehension in learning disabled and normally achieving students through self-questioning training. *Learning Disability Quarterly*, 5, 228–239.
- 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-Champaign, 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-Champaign, 210 Education Building, 1310 South Sixth, Champaign, IL 61820. She specializes in learning and instruction.

Received April 22, 1992

Revision received June 21, 1994

Accepted June 27, 1994