Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities

Annemarie Sullivan Palincsar and Ann L. Brown
Center for the Study of Reading
University of Illinois

Two instructional studies directed at the comprehension-fostering and comprehension-monitoring activities of seventh grade poor comprehenders are reported. The four study activities were summarizing (self-review), questioning, clarifying, and predicting. The training method was that of reciprocal teaching, where the tutor and students took turns leading a dialogue centered on pertinent features of the text. In Study 1, a comparison between the reciprocal teaching method and a second intervention modeled on typical classroom practice resulted in greater gains and maintenance over time for the reciprocal procedure. Reciprocal teaching, with an adult model guiding the student to interact with the text in more sophisticated ways, led to a significant improvement in the quality of the summaries and questions. It also led to sizable gains on criterion tests of comprehension, reliable maintenance over time, generalization to classroom comprehension tests, transfer to novel tasks that tapped the trained skills of summarizing, questioning, and clarifying, and improvement in standardized comprehension scores. Many of these results were replicated in Study 2. In contrast to Study 1, which was conducted by an experimenter, Study 2 examined group interventions conducted by volunteer teachers with their existing reading groups.

One of the most powerful tools of applied cognitive science is the training study (Chipman, Segal, & Glaser, in press). Guided by emergent theoretical analyses of the processes involved in a particular academic domain, researchers have designed cognitive skills training studies that have resulted in significant improvement in such areas as physics and mathematics problem solving

---

(Larkin, Heller, & Greeno, 1980), writing (Bereiter & Scardamalia, 1982) and in many of the multifaceted skills that underlie reading and studying (Baker & Brown, in press; Brown, Bransford, Ferrara, & Campione, 1983; Brown, Palincsar, & Armbruster, in press).

In this paper, we concentrate on improving students' ability to learn from texts. It is generally agreed that given reasonable facility with decoding, reading comprehension is the product of three main factors: (1) considerate texts, (2) the compatability of the reader's knowledge and text content, and (3) the active strategies the reader employs to enhance understanding and retention, and to circumvent comprehension failures.

Comprehension will be enhanced to the extent that the texts are well written, that is, they follow a familiar structure and their syntax, style, clarity of presentation, and coherence reach an acceptable level. Such texts have been called reader-friendly or considerate (Anderson & Armbruster, 1982).

Comprehension is also influenced by the extent of overlap between the readers' prior knowledge and the content of the text. Studies demonstrating the influence of schematic constructive processes in text comprehension are legion (Anderson, 1978; Mandler, 1983; Stein & Trabasso, 1982).

Then, there are strategies, processes for enhancing comprehension and overcoming comprehension failures. To illustrate the place of comprehension strategies in reading and studying, it helps to make the (over)simplified distinction between an "automatic" versus a strategic (or debugging) state (Brown, 1980; Collins & Smith, 1982). The well-practiced decoding and comprehension skills of expert readers permit them to proceed relatively automatically, until a triggering event alerts them to a comprehension failure. While the process is flowing smoothly, construction of meaning is very rapid; but when a comprehension failure is detected, readers must slow down and allot extra processing to the problem area. They must employ debugging devices or active strategies that take time and effort.

One commonly experienced triggering event is the realization that an expectation about the text has not been confirmed. Another triggering situation is when unfamiliar concepts are encountered too frequently for the reader to remain tolerant of his or her ignorance. Whatever the exact nature of the triggering event, mature readers react by slowing down their rate of processing, allocating time and effort to the task of clearing up the comprehension failure. In the process of disambiguation and clarification, they enter a deliberate, planful state that involves a variety of active processing strategies.

In a similar vein, when reading the text for retention (studying) as well as comprehension, the mature learner employs a whole variety of time-consuming activities to ensure that comprehension and retention are occurring:

> Studying actually requires a double or split mental focus. On the one hand, you
> need to be focused on the material itself. At the same time, however, you need

to be constantly checking to see that you are actually performing those mental operations that produce learning. In short, you need to monitor your mental processes while studying. (Locke, 1975, p. 126)

Studying involves the ability to plan strategies for learning and to test oneself concerning the effectiveness of any tactics one has called into service. When faced with the common task of understanding or committing material to memory, when time limitations or other restrictions impede unlimited study, learners must plan their time for most efficient results. Adequately dispensing the available study time involves an appreciation of which material is important, which material is not sufficiently well-known to risk a test, and methods of enhancing retention of both (Brown, Smiley, & Lawton, 1978).

Practiced readers proceed very differently when they are reading for pleasure or to obtain a quick impression of the gist, than when they are attempting to overcome a comprehension failure (debugging state), or when they are reading to meet strict criteria of understanding or retention (studying). In the first state, they read rapidly and, seemingly, effortlessly; but in the latter state, they proceed slowly and laboriously, calling into play a whole variety of learning and monitoring activities. This split mental focus, most clearly seen during studying, but also engaged in while reading in order to ensure against major comprehension failures, is the main focus of this paper.

Comprehension-Fostering and Comprehension-Monitoring Activities

In order to fully understand how a student learns from texts, one cannot ignore any of the four main factors of decoding fluency, considerate texts, compatible content, and strategic activity. But we have chosen to concentrate most extensively on the strategic component for two reasons. First, although it may indeed help novice learners to provide only considerate texts that focus on content knowledge that to a large extent they already possess, this surely is not the long-term goal of reading instruction. Mature readers can come to grips with a variety of inconsiderate texts, that is, those that creatively violate accepted structure or those that are just plain poorly written. Mature readers also read to learn, that is, they read to obtain content knowledge that they do not yet possess. In order to deal with unfamiliar content often presented in less than hospitable forms (some textbooks, for example), the reader needs to employ strategies.

The second reason we have chosen to concentrate on comprehension-fostering strategies is that they comprise a set of knowledge-extending activities that apply in a wide range of situations other than reading; these are the basic skills of argument. Mature learners question and elaborate their own knowledge and the content of the text, testing their degree of understanding

by thinking of counter-examples and testing possible generalizations, by attempting to apply their new-found knowledge, and by a variety of debugging ploys that force them to correct their misunderstandings (Collins & Stevens, 1982). In the face of difficulty, they use a variety of fix-up strategies, from simply rereading difficult segments to constructing elaborate scenarios to explain incompatible or unexpected events. These knowledge-extending activities are as applicable when one is listening or debating as when one is reading. As there is ample evidence that such activities prove particularly troublesome for the young and the academically weak (Brown, Armbruster, & Baker, in press; Brown & Campione, 1981, in press; Brown & Palincsar, 1982), attempts to instruct these activities seem particularly worthwhile.

One daunting problem for those who would engage in the explicit instruction of comprehension skills is that there are so many putative strategies, descriptions of which are often quite vague. There is, however, considerable agreement concerning what the most important underlying activities might be. In a review of both the traditional reading education literature and recent theoretical treatments of the problem, we found that six functions were common to all (Brown, Palinscar, & Armbruster, in press): (1) understanding the purposes of reading, both explicit and implicit; (2) activating relevant background knowledge; (3) allocating attention so that concentration can be focused on the major content at the expense of trivia; (4) critical evaluation of content for internal consistency, and compatibility with prior knowledge and common sense; (5) monitoring ongoing activities to see if comprehension is occurring, by engaging in such activities as periodic review and self-interrogation; and (6) drawing and testing inferences of many kinds, including interpretations, predictions, and conclusions.

For the purposes of instruction, we selected four concrete activities that could be engaged in by novice learners and that would embody the overlapping functions contained in points 1 through 6 above. These were *summarizing* (self-review), *questioning, clarifying,* and *predicting.* By asking students to summarize a section of text, one is simultaneously requesting that they allocate attention to the major content (3) and that they check to see if they have understood it (5). In requesting that students compose questions on the content, one is also asking for a concentration on main ideas (3) and a check of the current state of understanding (5). Asking students to clarify requires that they engage in critical evaluation as they read (4), and asking them to make predictions concerning future content involves them in drawing and testing inferences (6). All four activities involve activation of relevant background knowledge (2). In addition to the four strategies, points 1 and 2 above were addressed by embedding the instruction in the context of reading for the clear purpose of answering questions on the text, and by discussing relevant background knowledge at the start of each instructional period (Au, 1979).

In summary, these four activities were selected because they provide a dual function, that of enhancing comprehension and at the same time affording an opportunity for the student to check whether it is occurring. That is, they can be both *comprehension-fostering* and *comprehension-monitoring* activities if properly used. Self-directed summarization is an excellent comprehension-monitoring technique (Brown & Day, 1983; Brown, Day, & Jones, 1983; Day, 1980; Linden & Wittrock, 1981). Monitoring one's progress while reading, to test whether one can pinpoint and retain important material, provides a check that comprehension is progressing smoothly. If the reader cannot produce an adequate synopsis of what she is reading, this is a clear sign that comprehension is *not* proceeding smoothly and that remedial action is called for. Similarly, self-directed questioning concerning the meaning of text content leads students to a more active monitoring of their own comprehension (Andre & Anderson, 1978–79). Thus, closing one's eyes (metaphorically) and attempting to state the gist of what one has read, and asking questions of *clarification, interpretation,* and *prediction* are activities that both improve comprehension and permit students to monitor their own understanding.

These are also the kinds of active and aggressive interactions with texts that poor readers do not engage in readily. For example, although quite young children understand the essential features of a summary (Brown, Day, & Jones, 1983; Johnson, 1983), students cannot adequately summarize a typical fifth-grade academic text until well into the high school years; academically delayed students, that is, remedial readers, have not mastered this ability by the time they reach (junior) college (Brown & Day, 1983). Documentation of students' difficulties generating questions on what they are reading is legion, and again the problem is particularly acute for the slower student (Andre & Anderson, 1978–79). Similarly, there is considerable recent work documenting the fact that young and poor readers have difficulty evaluating texts for clarity, internal consistency, or compatibility with known facts (Baker, 1983; Markman, 1981). And the ability to interpret what is happening now and what will happen next is also slow to develop in academically marginal students (Baker & Brown, in press; Collins & Smith, 1982). The need for explicit instruction in comprehension-enhancing activities is particularly acute in the slow-learning student (Brown, Bransford, Ferrara, & Campione, 1983; Brown & Palincsar, 1982; Franks, Vye, Auble, Mezynski, Perfetto, Bransford, Stein, & Littlefield, 1982).

Therefore, in this series of studies, we decided to train the four activities of self-directed summarizing (review), questioning, clarifying, and predicting, embedding them in the context of a dialogue between student and teacher that took place during the actual task of reading with the clear goal of deriving meaning from the text. Each "separate" activity was used in re-

sponse to a concrete problem of text comprehension. Summarizing was modeled as an activity of self-review; it was engaged in to state to the teacher or the group what had just happened in the text and as a self-test that the content had been understood. If an adequate synopsis could not be reached, this fact was regarded *not* as a failure to perform a particular decontextualized skill, but as an important source of information that comprehension was not proceeding as it should, and remedial action (such as rereading or clarifying) was needed. Similarly, questioning was not practiced as an isolated activity, but as a continuing goal of the whole enterprise — what main idea question would a teacher or test ask about that section of the text? Clarifying occurred only if there were confusions either in the text (unclear referent, etc.) or in the student's interpretation of the text. And prediction was attempted if the students or teachers recognized any cues that served to herald forthcoming material. In short, all of the activities were undertaken in the context of actually reading with the goal of understanding and remembering the text content.

Interactive Training

We turn now to the instructional mode, how to teach the activities. One main concern was to try to avoid a common problem with traditional training studies, the outcomes of which have been somewhat discouraging. Although improvement on a particular skill in isolation has been reported, this improvement is often slight and fleeting, and there is very little evidence of transfer. Maintenance over time, generalization across settings, and transfer within conceptual domains are rarely found. The more difficulties the learner experiences initially, the more fleeting and bounded are the effects of training (Borkowski & Cavanaugh, 1979; Brown, 1978; Brown & Campione, 1978, 1981, in press; Brown, Palincsar, & Purcell, in press; Meichenbaum & Asarnow, 1978).

One explanation of this typical pattern has been the general diagnosis of the trainee as a relatively passive participant who responds to instruction but does not fully understand the activities she has been induced to use (Brown, 1978). Training studies that have (1) forced the student to be active, (2) provided feedback in the utility of that action, and (3) provided instruction in why, when, and where such activities should be applied, have been much more successful in inducing transfer (Brown, Day, & Jones, 1983). We needed, then, a mode of instruction that would ensure that the students would participate at whatever level they were capable, that would enable them to witness the success of such activities, and that would situate the strategies in an actual reading context where the goal of the activities would be transparent.

We designed an intervention that, in addition to meeting the above criteria, was based on the notions of expert scaffolding and proleptic teaching (Rogoff & Gardner, in press; Wertsch & Stone, 1979). Proleptic means "in

anticipation of competence," and in the context of instruction refers to situations where a novice is encouraged to participate in a group activity before she is able to perform unaided, the social context supporting the individual's efforts. In these teaching situations, a novice carries out simple aspects of the task while observing and learning from an expert, who serves as a model for higher level involvement. Work on proleptic instruction has been influenced by Vygotsky's (1978) developmental theory. Vygotsky believed that a great deal of development was mediated by expert scaffolding. Children first experience a particular set of cognitive activities in the presence of experts, and only gradually come to perform these functions by themselves. First, an expert (parent, teacher, mastercraftsman, etc.) guides the child's activity, doing most of the cognitive work herself. The child participates first as a spectator, then as a novice responsible for very little of the actual work. As the child becomes more experienced and capable of performing more complex aspects of the task, aspects that she has seen modeled by adults time and time again, the adult gradually cedes her greater responsibility. The adult and child come to share the cognitive work, with the child taking initiative and the adult correcting and guiding where she falters. Finally, the adult allows the child to take over the major thinking role and adopts the stance of a supportive and sympathetic audience. Initially, the supportive other acts as the model, critic, and interrogator, leading the child to use more powerful strategies and to apply them more widely. In time, the interrogative, critical role is adopted by the child, who becomes able to fulfill some of these functions for herself via self-regulation and self-interrogation. Mature learners are capable of providing the interrogative critical role for themselves (Binet, 1909; Brown, in press).

Within these systems of tutelage, the child learns about the task at his own rate, in the presence of experts, participating only at a level he is capable of fulfilling—or a little beyond, thereby presenting a comfortable challenge. The distance between the level of performance a child can reach unaided and the level of participation he can accomplish when guided by another more knowledgeable than he is the child's zone of proximal development (Vygotsky, 1978), or "region of sensitivity to instruction" (Wood & Middleton, 1975). For a particular child in a particular domain, this zone may be quite small, the interpretation being that the child is not yet ready to participate at a more mature level than his unaided performance would indicate. For another child in that domain, the zone of proximal development can be quite dramatically large, indicating that with aid, sometimes minimal aid at that, he can participate much more fully and maturely in the activity than one might suppose on the basis of only his unaided performance. In traditional terms, these are notions of readiness. What is distinct about Vygotsky's theory is the important role attributed to the social context and expert scaffolding.

It has been argued that children originally learn to apply comprehension-fostering activities to text in proleptic teaching interactions very similar to the classic variety described by Vygotsky. For example, descriptions of parent–child picture book rituals (Ninio & Bruner, 1978), as well as teacher–pupil instruction in "good" reading groups (Allington, 1980; Collins, 1980; McDermott, 1978) are excellent examples of an expert guiding the child as he attempts to apply comprehension-fostering activities to texts. It has also been argued that, for a variety of reasons, children at risk for academic failures are less likely to be exposed to ideal interactive learning situations; they are said not to occur as often in disadvantaged homes (Heath, 1981), and children in "poor" reading groups receive more attention to their decoding skills at the expense of their comprehension-fostering activities (Collins, 1980). Lack of experience applying basic thinking skills to the task of reading may be an important contributing factor to the slow development of such skills in many children (Meichenbaum & Asarnow, 1978).

Following this argument, we attempted to design an intervention that would mimic, as closely as possible, naturally occurring guided learning interactions in which the teacher could both model appropriate comprehension-fostering activities and at the same time guide the child to participate at an ever-increasing level of competence.

Pilot Study

In a pilot study (Brown & Palincsar, 1982), we developed a procedure, *reciprocal teaching*, where teacher and student took turns leading a dialogue concerning sections of a text. The procedure was similar to, but more extensive than, the reciprocal questioning intervention used with some success by Manzo (1968). In addition to reciprocal questioning, the teacher and students took turns generating summaries and predictions and in clarifying misleading or complex sections of the text. Initially, the teacher modeled the key activities of summarizing (self-review), questioning (making up a question on the main idea), clarifying, and predicting. The teacher thereby demonstrated the appropriate activities; the students were encouraged to participate at whatever level they could. The teacher could then provide guidance and feedback at the appropriate level for each student.

The basic procedure was that an adult teacher, working individually with a seventh-grade poor reader, assigned a segment of the passage to be read and either indicated that it was her turn to be the teacher or assigned the student to teach that segment. The adult teacher and the student then read the assigned segment silently. After reading the text, the teacher (student or adult) for that segment asked a question that a teacher or test might ask on the segment, summarized the content, discussed and clarified any difficulties, and finally made a prediction about future content. All of these activities were

embedded in as natural a dialogue as possible, with the teacher and student giving feedback to each other.

Initially, the adult teacher modeled the activities, but the students had great difficulty assuming the role of dialogue leader when their turn came. The adult teacher was sometimes forced to construct paraphrases and questions for the students to mimic. In this initial phase, the adult teacher was modeling effective comprehension-monitoring strategies, but the students were relatively passive observers.

Gradually, the students became much more capable of assuming their role as dialogue leader and by the end of ten sessions were providing paraphrases and questions of some sophistication. For example, in the initial sessions, half of the questions produced by the students were judged as nonquestions or as needing clarification; however, by the end of the sessions, unclear questions had dropped out and were replaced with questions focusing on the main idea of each text segment. A similar improvement was found for summary statements. At the beginning of the sessions, only a few summary statements captured main ideas, whereas at the end, the majority of the statements were so classified (Brown & Palincsar, 1982; Palincsar, 1981).

Each day, before (baseline), during, and after (maintenance) training, the students took an unassisted assessment, where they read a novel passage and answered ten comprehension questions on it from memory. From their baseline performance of 15% correct, they improved during training to accuracy levels of 85%, levels they maintained when the intervention was terminated. Even after a 6-month delay, the students averaged 60% correct without help, and it took only 1 day of renewed reciprocal teaching to return them to the 85% level achieved during training. Remember that these scores were obtained on *privately read* assessment passages, that is, different texts that the students read independently after their interaction with the instructor. What was learned during the instructional sequence was used independently by the learners.

Given the success of the pilot work, we undertook the current more extensive studies. In Study 1, the performance of six students receiving the reciprocal teaching intervention was compared with that of matched students assigned to a variety of control groups. The teacher in Study 1 was the first author; in contrast, in Study 2, the teachers were volunteer reading teachers who conducted the reciprocal teaching intervention with their pre-existing reading groups. In addition, in both Study 1 and Study 2, we included multiple and stringent criteria of success of the intervention. These included (1) reliable improvement on the training task; (2) independent evidence of improvement in the strategies trained; (3) improvement in the students' independent reading of novel passages; (4) durability of the effect of training; (5) generalization of the effects across settings, notably to the classroom; and (6) transfer to novel tasks that demand the same underlying processes but

differ in surface structure from the training vehicle. Previous cognitive skills training studies have included no more than two of these criteria; indeed, the majority have included only the first (Brown, Palinscar, & Armbruster, in press; Campione & Armbruster, in press). We were particularly concerned with including transfer measures to assess what, if any, general improvement followed the intervention.

STUDY 1

Method

Subjects.   Thirty-seven seventh-grade students took part in this study. Twenty-four of the students had reading problems, 13 did not. The 24 poor readers were selected from a larger group nominated by their teachers as adequate decoders but poor comprehenders. Although receiving attention as poor readers, they were *not* labeled as learning disabled or mentally retarded. To test the accuracy of the teachers' diagnosis, we administered our own tests of decoding and comprehension to the poor readers. To meet the criteria of adequate decoding, the students were required to read grade-appropriate texts at a rate of at least 80 wpm with two or fewer errors. This criterion was established by Lovitt and Hansen (1976a) as the *minimum* acceptable decoding fluency for instructional purposes. Next, their putative comprehension difficulties were determined by establishing that they met two criteria: (1) standardized reading comprehension scores at least 2 years below grade level; and (2) baseline performance of below 40% correct on the experimental task.

Initially, six students meeting these criteria were assigned to the reciprocal teaching condition and a further six to an untreated control group (reported in Palinscar, 1982). Subsequently (6 months later), a further 12 subjects meeting these criteria were assigned to two additional groups, one involving an alternative intervention, locating information, and another involving practice with the daily assessment passages but no intervention. For convenience, we will discuss all four groups together.

The total of 24 subjects meeting the minimal decoding standards but failing on the three measures of comprehension, were, then, divided into four groups with six subjects per group. Two of the groups received instruction. Group 1 received *reciprocal teaching* and Group 2 received *locating information,* both to be described below. In addition, there were two control groups. Group 3, *test only,* received all of the daily assessment tests but no intervention. The final group, *control* (Group 4), received only the pretests (baseline) and posttests (maintenance), but remained in their normal classrooms for reading instruction. The descriptive statistics for these 24 students are shown in Table 1. Note that the groups are quite closely matched, scoring approxi-

mately 2 1/2 years delayed on standardized reading comprehension scores and showing oral reading rates of approximately 100 words per minute, with less than two errors per minute. The only difference between the groups is the baseline comprehension assessment score, which was lower for the reciprocal teaching group than for the remaining groups.

In addition, we administered individual IQ tests to the six students in the reciprocal teaching group. These IQs ranged from 67 to 99, with a mean of 83. Time and the availability of subjects did not permit us to test the remaining students. However, their vocabulary scores (shown in Table 1), are very similar to those of the reciprocal teaching group, on average 2 years delayed. Together, the available IQ scores and vocabulary measures suggest that our subjects were performing at the lower end of the normal distribution of academic intelligence.

In order to provide some normative data on the tests to be used in the study, 13 seventh-grade volunteers contributed data to the baseline and maintenance sessions, along with performing on the transfer posttests. They were recruited from a summer Bible class, and came to the laboratory during their summer vacation. They were all seventh graders with chronological ages appropriate to their grade placement. None of them had skipped a grade or been held back during their academic career. None of them reported any history of academic delay, their average grade in classroom exercises was B +, and they reported being at grade level on standardized tests of reading (parents and students self-report). Unfortunately, we were not able to verify their standardized test status. We will refer to them throughout as average seventh grade readers.

Materials

A total of 13 passages were available for training, averaging 1500 words each in length. They were selected from the following reading series: *Reading Unlimited* (Aaron, et al., 1976; *Keys to Reading* (Matteoni, Lane, Ducher, & Burns, 1980); *Adventures for Readers* (Nieman & Safier, 1979); *Reading 720* (Clymer, 1976); *Corrective Reading Decoding* (Engelman, Becker, Hanner, & Johnson, 1978); *Serendipity* (Durr, Pescosolido, & Toetter, 1974). All of the passages were expository and represented a range of topics including: poisonous snakes, solar energy, the Inca civilization, lightning, and carnivorous plants. The passages were selected after determining that they conformed to a seventh-grade level according to the Fry Readability Formula.

In addition to the training passages, a total of 45 shorter assessment passages were selected from the same sources as the training material. The assessment passages were expository, written at a seventh-grade readability level (according to the Fry formula), and ranged in length from 400 to 475 words.

TABLE 1
Descriptive Statistics of the Seventh-Grade Students in Study 1

	Subject #	Comprehension		Vocabulary		Oral Reading, wpm		Baseline
		GE[1]	Delay	GE[1]	Delay	Correct	Incorrect	% Correct
Group 1:	1	4,2	−2,10	5,5	−1,7	99.9	1.7	30
Reciprocal	2	3,6	−3,6	4,2	−2,10	110.0	1.8	20
Teaching	3	5,3	−1,9	6,5	−0,7	134.0	0.8	20
	4	6,4	−0,8	4,9	−2,3	131.0	0.7	40
	5	3,9	−3,3	5,2	−1,10	99.0	1.1	0
	6	3,1	−3,11	5,2	−1,10	121.0	0.7	30
	$\overline{X}$	4,4	−2,7	5,3	−1,8	115.8	1.1	23.3
Group 2:	7	3,3	−3,9	—	—	89.6	0.8	50
Locating	8	4,4	−2,8	4,3	−2,9	119.0	0.0	50
Information	9	4,2	−2,10	4,2	−2,10	95.2	2.9	20
	10	5,3	−1,9	4,6	−2,6	95.8	2.3	30
	11	4,4	−2,8	3,9	−3,3	111.4	0.9	40
	12	3,6	−3,6	6,3	−0,9	105.7	0.9	20
	$\overline{X}$	4,2	−2,9	4,7	−2,4	102.8	1.3	35.0

Group 3:	13	4,4	−2,8	4,3	−2,9	79.5	2.5	20
Test Only	14	4,7	−2,5	4,6	−2,6	117.0	2.0	20
	15	5,3	−1,9	4,6	−2,6	104.1	0.9	40
	16	4,7	−2,5	4,6	−2,6	100.0	1.7	50
	17	5,6	−1,6	6,0	−1,0	101.7	2.3	40
	18	5,4	−1,8	5,0	−2,0	104.8	1.8	50
	$\overline{X}$	5,0	−2,1	4,9	−2,2	101.2	1.9	36.7
Group 4:	19	5,3	−1,9	6,4	−0,8	89.6	1.3	30
Control	20	3,7	−3,5	3,2	−3,10	141.6	1.2	20
	21	4,3	−2,9	6,5	−0,7	152.6	2.6	50
	22	3,2	−3,10	6.0	−1,0	168.6	1.4	40
	23	4,3	−2,9	8,7	+1,7	127.5	0	30
	24	4,2	−2,10	3,9	−3,3	124.2	1.1	30
	$\overline{X}$	4,2	−2,9	5,8	−1,3	130.9	1.3	33.3

[1]GE = Grade Equivalent, Gates-MacGinitie Test. Grade equivalent score = grade plus additional months, i.e., 4,2 = fourth grade plus 2 months.

The 10 comprehension questions per assessment passage were constructed by the first author using the Pearson and Johnson (1978) classification of: (1) *text explicit,* where the answer was explicitly mentioned in the text; (2) *text implicit,* where the answer needed to be inferred by combining across adjacent segments of text; and (3) *script implicit,* where the answer must be arrived at by considering text in relation to prior knowledge concerning the topic in question. Two independent raters (English teachers) agreed to the classification of question types for each passage. Of the 450 original questions, the raters disagreed on the classification of only 15. These were rewritten until agreement was reached.

Procedure

Control Groups

There were two control groups who received no intervention and two treatment groups, reciprocal teaching and locating information. Consider first the two control groups, Test Only and Control. The first control group, referred to as Test Only (Group 3), consisted of six students seen daily in groups of two. They received no intervention but read and answered questions on the daily assessment passages. They also took the baseline, maintenance, and pretests and posttests. The second control group (Group 4, Control) took only the baseline, maintenance, and pretests and posttests; they did not receive any daily practice with the assessment passages. Instead they remained in their normal reading class for that period.

Treatment Groups

Instead of their normal reading group, the students in the treatment groups received either *reciprocal teaching* (Group 1) or *locating information* (Group 2). For both treatment groups, there were four phases to the study: (1) *variable baseline,* consisting of 4 days for the first two students of each group, 6 days for the second two, and 8 days for the remaining two students in each group, (2) *intervention,* consisting of approximately 20 days; (3) *maintenance,* consisting of 5 days of testing at the termination of training; and (4) *long-term follow-up,* 3 days of testing that took place 8 weeks later. All students in the Reciprocal Teaching, Locating Information, and Test Only groups were apprised of their progress on a daily basis. They were shown graphs depicting the percentage correct for the previous day's assessment, and a weekly cumulative record. The students were given daily assessment passages on which they answered 10 comprehension questions, and this was all that occurred for the treatment groups during baseline, maintenance, and follow-up.

Reciprocal Teaching.   During intervention, the Reciprocal Teaching group received the intervention outlined in the description of the pilot study.

Briefly, the six students were assigned to three groups of two students who worked together with the adult teacher. Each day the teacher would introduce the passage with a brief discussion that was intended to activate the student's prior knowledge (Au, 1979). If at the beginning of instruction the passage were new, the investigator called the student's attention to the title and asked for predictions based on the title. If the passage were partially completed (portions having been covered on a previous day), the investigator asked the students to state the topic of the text and several important points already covered in the passage.

The adult instructor then assigned a teacher for the first segment (usually a paragraph) and the group read the segment silently. Then the teacher for that segment proceeded first to ask a question, then to summarize, and to offer a prediction or ask for a clarification when appropriate.

The adult teacher provided the guidance necessary for the student teacher to complete the preceding activities through a variety of techniques: *prompting*, "What question did you think a teacher might ask?"; *instruction*, "Remember, a summary is a shortened version, it doesn't include detail"; and *modifying the activity*, "If you're having a hard time thinking of a question, why don't you summarize first?"

The adult teacher also provided praise and feedback specific to the student's participation: "You asked that question well; it was very clear what information you wanted." "Excellent prediction; let's see if you're right." "That was interesting information. It was information that I would call detail in the passage. Can you find the most important information?" After this type of feedback, the adult teacher modeled any activity which continued to need improvement: "A question I would have asked would be..." "I would summarize by saying..." "Did you find this statement unclear?" Interactive sessions lasted approximately 30 minutes. All dialogues were tape-recorded so that qualitative changes could be assessed.

Throughout the interventions, the students were explicitly told that these activities were general strategies to help them understand better as they read, and that they should try to do something like this when they read silently. It was pointed out that being able to say in your own words what one has just read, and being able to guess what the questions will be on a test, are sure ways of testing oneself to see if one has understood.

Locating Information. Students in the Locating Information group were treated identically, that is, seen in groups of two and given baseline, intervention, and maintenance testing. They took the same assessment passages each day and were shown graphs of their previous day's performance and a weekly cumulative record. Only the intervention varied.

The locating information treatment was modeled on a procedure commonly used by remedial reading teachers to help students cope with answering questions on texts they have just read (Bartlett, 1979; Lovitt & Hansen,

1976b; Raphael, 1980). There is empirical support for the success of this intervention. Lovitt and Hansen (1976b) employed a locating information condition when comparing the effects of drill and reinforcement on the decoding and comprehension skills of reading impaired students. While decoding fluency rates increased with reinforcement alone, drill — having students correct their answers to comprehension questions while referring to the text — was necessary to improve accuracy with comprehension questions. Similarly, Raphael, Wonnocat, and Pearson (1983) have conducted a series of studies in which they successfully taught students specific strategies for locating information to respond to the same question types used here. Finally, in the pilot study (Brown & Palincsar, 1982; for full details, see Palincsar, 1982), locating information did result in reliable improvement, although it was not as extensive or durable as that resulting from reciprocal teaching.

The locating information procedure involves demonstration and practice in test taking. The students are shown where in the text the answer to a text-explicit question can be found. The teacher demonstrates how to combine across separate sections of text to answer a text-implicit question, and the students are shown how to use their prior knowledge to answer a script-implicit question.

After reading the assessment passages independently (help was given if they couldn't understand anything), the student and teacher answered the questions together with the text in front of them. The student attempted to answer. The teacher praised correct responses; following incorrect responses, she guided the student back into the passage to the appropriate paragraph where the answer could be found. If necessary, the line(s) where the answer could be found was given, as well as prompts to help the students find the answer. During the procedure, the students were being taught that the answers to the questions could be found with a little work with the text and their prior knowledge, a proposition that they greeted with some surprise! Answers to questions were discussed and mutually agreed upon by student and investigator. On the following assessment passages, the students worked unaided and answered questions from memory, just as in the Reciprocal Teaching group.

Generalization and transfer.   For two of the groups (the original two run), Reciprocal Teaching and the Control, generalization and transfer tests were taken before, during, and after the intervention. We were not able to collect these measures on the Locating Information and Test Only groups, who were run one term later.

Generalization probes in the classroom were taken five times during the course of the study. The first probe occurred during baseline; the second and third occurred during the first and second half of the interventions, respectively. The fourth probe occurred during maintenance; the fifth and final

probe occurred during follow-up, 8 weeks after the termination of training. The tests consisted of passages followed by 10 comprehension questions similar to the daily assessments. However, the generalization probes occurred in regular social studies and science classes, with no notification given to the students that these tests were part of the study. The entire class took the tests as part of their regular work in the class. The passages used in the generalization probes were also taken from the books actually in use in the classes: *World Geography* (Backler & Lazarus, 1980); and *Life Science* (Heimler & Lockard, 1977). The text segments were selected in consultation with the regular classroom teachers and featured materials not yet introduced to the students, so that reading comprehension was not confounded with prior instruction. Each segment was approximately 450 words long and written at a seventh-grade readability level (Fry, 1977). The 10 comprehension questions met the same criteria as those described for the daily comprehension questions. The teachers approved the questions as "the type one should ask students," although they expressed considerable doubt that these particular students (poor comprehenders) could handle them.

Transfer tests. These were selected because we believed that they tapped the skills taught during the reciprocal teaching, and, pragmatically, because a considerable body of prior work has established "normal" levels of performance for seventh graders. They were given a pretest-posttest format, prior to and upon the termination of the study.

Although there is considerable controversy concerning the definition of transfer (Brown, 1978; Brown & Campione, 1981, in press) versus maintenance of a trained skill, we adopted the procedure of selecting tasks that, although clearly demanding the application of the trained skills, differed considerably in surface structure from the intervention or the daily assessments. Thus, although the same skills could fruitfully be applied, the actual task formats in training and transfer were quite distinct.

Two of the four transfer tests were measures of the two most frequently engaged in activities during the reciprocal teaching sessions, *summarizing* (Brown & Day, 1983) and *predicting questions* that might be asked concerning each segment of text (Wong & Jones, 1981). In addition, two other tests were used as measures of general comprehension monitoring, *detecting incongruities* (Harris, Kruithof, Terwogt, & Visser, 1981; Markman, 1979) and *rating importance* of segments of narratives (Brown & Smiley, 1977).

Summarizing. The procedure here was a simplified version of that used by Brown and Day (1983) to examine students' use of various macrorules (Kintsch & vanDijk, 1978) for considering texts. The students were given fifth-grade expository texts (from Brown & Day, 1983; and Day, 1980), two prior to and two after training. They were asked to indicate where

any of five main condensing rules could be applied. The five rules were: (1) *deletion* of trivia; (2) *deletion* of redundancy; (3) *superordination,* where a list of exemplars was replaced with a superordinate term; (4) *selection* of a topic sentence to serve as a scaffolding of the summary, and (5) *invention* of a topic sentence for a paragraph where one was not explicitly stated (see Brown & Day, 1983, for details). Brown and Day found that unaided seventh graders had great difficulty using any but the deletion rules; therefore, we simplified the task, and (1) told students what the five rules were and wrote them on the board with examples and (2) asked students to apply these rules to the two written texts in front of them. For example, the students deleted (crossed out with a red pencil) all sections they judged to be trivia, and (with a blue pencil) all sections they judged to be redundant. They also wrote in a superordinate term for any lists. They underlined appropriate topic sentences and wrote in inventions when they judged that a paragraph lacked an explicit statement of topic.

Predicting questions.   The ability to generate important and clear questions was a skill which received considerable focus during reciprocal teaching. The following measure was included to assess the accuracy with which students could identify and construct "teacher-like" questions. The students were given four randomly assigned passages, two prior and two following the study. They were asked to write 10 questions a classroom teacher might ask if testing the student's knowledge of the passage. The passages were taken from material written at a seventh-grade level (Fry, 1977).

Detecting incongruities.   One popular index of comprehension monitoring is the ability to detect errors or anomalies in text (Markman, 1979). We adopted a procedure similar to that used by Harris et al., (1981). Eight stories were prepared: For six of the stories, one line was anomalous with the title; the remaining two stories made sense. Four of the stories, three containing anomalous lines, were presented on the pretest and the remainder on the posttest. Each story was presented line-by-line on an Apple II computer. The students were told to read each line and say "yes" if the line made sense in the story, or "no" if the line didn't make sense. Their responses were timed. After reading and appraising each line of the story, the entire story was returned to the screen. To avoid contriving a situation in which students began to have expectations regarding the passages, a very general probe was used. For any story evaluated by the student as all correct, the examiner said, "Here is the entire story. You have decided that every line made sense. Is that correct?" For any line evaluated by the student as incorrect, the investigator pointed to the line(s) and said, "Can you tell me why this line doesn't make sense in the story?"

Rating importance.   Four narrative passages (from Brown & Smiley, 1978) were selected as measures of sensitivity to main idea and detail information. They were randomly administered to each student, two before and two after the intervention. The students were asked to first read the intact stories. They were then presented the stories with each data unit typed on a separate line. The students were told that the stories were to be rewritten for the purpose of fitting them into tiny doll house books and that they were to choose only the most important lines. It was explained that they were first to delete N lines (1/4 of the text) by crossing out the least important with a blue pencil. They were then asked to eliminate the next N unnecessary lines, using a green pencil. Finally, they were asked to cross out another N lines in red, leaving only the most important lines for inclusion in the doll house books.

RESULTS AND DISCUSSION

Dialogue Changes: Reciprocal Teaching

All dialogues of the Reciprocal Teaching groups were recorded, transcribed, and scored by independent raters. As we had previously found in the pilot study, unclear questions and detailed summaries predominated in the early sessions, while main idea questions and summaries were most common in the later sessions. Main idea questions increased from 54% to 70% of the total. The quality of the summary statements also improved, with incorrect and incomplete statements declining from 19% to 10% and detail summaries declining from 29% to 4%. These declines in inadequate summaries were accompanied by a significant increase in main idea summaries from 52% to 85% of the total ($z = 2.86$, $p < .002$). Examples of inadequate questioning and summarizing occurring in the early sessions, and improved versions occurring in the later sessions, can be found in Table 2 and Table 3.

With repeated interaction with an adult model performing appropriate questioning and paraphrasing activities, the students became able to perform these functions on their own. Over time, the students' questions became more like the tutor's, being classified as *inventions,* that is, questions and summaries of gist in one's own words, rather than *selections,* repetitions of words actually occurring in the text (Brown & Day, 1983). For example, an early occurring form of question would be to take verbatim from the text, "plans are being made to use nuclear power," and append the question with the inflection, "for what?" Later forms of questioning were most likely to be paraphrases of the gist in the students' own words. For example, reading a passage about fossils, one student posed the following question: "When an animal dies, certain parts decay, but what parts are saved?" This question

TABLE 2
Examples of Student Generated Questions During Reciprocal Teaching

First half of training

    Questions requiring clarification (and ideal questions regarding the same materials and ideas)

What was uh, some kings were uh, about the kings? (Why is it that kings did not always make the best judges?)

What were some of the people? (What kinds of people can serve on a jury?)

What was the Manaus built for? Wait a minute. What was the Manaus built for, what certain kind of thing? Wait a minute. O.K. What was the Manaus tree built for? (Why was the city of Manaus built?)

What does it keep the ground? (What effect does snow have on the ground?)

What are the Chinese people doing today, like ... What are they doing? (Why are the Chinese people rewriting their alphabet today?)

There's you know, like a few answers in here and one of my questions is, uh, anything that burns and explodes can be fast enough to ... See, they got names in here. O.K.? (Name some explosives.)

In Africa, India, and the Southern Islands where the sun shines what happens to the people? You know, like ...? (Why do people who live in Africa, India and the Southern Islands have dark skin?)

    Detail questions (and ideal main idea questions on the same material)

Do water moccasins have a joint in their eyes—nostrils—which they are able to—they are sensitive to? (How is a water moccasin able to tell that a warm blooded animal is near?)

How many backbones can a snake have? (What is unusual about the vertebrae of a snake?)

The western corral is only a blank, blank, and blank long. (Name an important characteristic of the western coral snake.)

How do camels spread out in a line? (What is a caravan?)

Do most of the people when they get camels, save them or do they try to get rid of them? (In what ways are camels useful to people who live in the desert?)

What color is the guard's uniform? (What would you see if you were visiting Buckingham Palace?)

Second half of training

    Main idea questions

Where does lightning get its charge?

When a larva reaches its full size, what does it become?

Where does the pitcher plant grow and why does it grow there?

How do you know that the winds are really strong in Antarctica?

Why would scientists go to Antarctica?

What does hemoglobin do?

Why don't people live in the desert?

Why are the grasslands of Australia ideal for grazing?

What did these people (the Chinese) invent?

What are three main problems with all submarines?

136

TABLE 3
Examples of Student Generated Summaries During Reciprocal Teaching

First half of training

  Incomplete summaries (and suggested appropriate summaries)

They talk about it was the richest island; but it didn't have something, O.K., it was the
    richest island but didn't have everything. They didn't have something. (Although this
    was a very rich land, no people lived there.)
If you pick a cherry branch in the winter you will have luck hoping they will bloom early.
    (If you pick a cherry branch in the winter, you will have no luck with it blooming.)
And uranium can be making explosion that equals a skyscraper. (A small amount of
    uranium can cause an explosion as great as a skyscraper full of dynamite.)
What camels do ... (Camels are invaluable to the people of the desert.)
Snakes can move forward and backward and they have a rubbery something. (A snake's
    flexibility is due to its unusual backbone, which consists of many vertebrae connected
    by loose rubbery tissue.)

  Detailed summaries (and suggested appropriate summaries)

It's about the coral snake. It's about how long the coral snake is and the western coral
    snake. And the next, they're harmful, except for fangs are too short to go through
    heavy clothes or shoes. (This paragraph describes the physical characteristics of the
    common and western coral snakes.)
I can summarize. It's about like the full grown spider, when it wants to travel it has to use
    all eight legs and the much younger ones have to use posts, bushes, and fence posts,
    and posts, and leaf stems. And it tells you like when the wind blows how the young
    spiders hold onto their streamers. (This paragraph describes two methods by which
    spiders travel from one location to another.)

Second half of training

  Main idea summaries

It's about how Hawaii is a chain of undersea volcanoes and about the myth of how they
    were formed.
This is about lightning flashes and that it's caused by two different clouds.
This one is mainly about the stages of the butterfly.
This is all about how the venus fly trap lures and captures its victim.
I would summarize it by saying that this tells about the red blood cells and the hemoglobin
    that carries gasses around.
This is about how crows communicate with each other.
It tells us about the two kinds of camels, what they are like and where they live.

was constructed by integrating information presented across several
sentences.

   Close inspection of the dialogues revealed repeated examples of guided
learning, that is, where the teacher provided modeling, feedback, and prac-
tice to students at a level that appeared to match the student's current need.
As students became better able to perform some aspects of the task, the

TABLE 4
The Acquisition of Question-Asking by a Weak Seventh-Grade Student (Charles)

Day 1:

TEXT: The water moccasin, somewhat longer than the copperhead, is found in the southeastern states. It lives in swampy regions. It belongs, as do also the copperhead and the rattlesnakes, to a group of poisonous snakes called pit vipers. They have pits between their eyes and their nostrils which, because they are sensitive to heat, help the snakes tell when they are near a warm-blooded animal. Another name for the water moccasin is "cottonmouth." This name comes from the white lining of the snake's mouth.

1. C: What is found in the southeastern snakes, also the copperhead, rattlesnakes, vipers—they have. I'm not doing this right.
2. T: All right. Do you want to know about the pit vipers?
3. C: Yeah.
4. T: What would be a good question about the pit vipers that starts with the word "why?"
5. C: (No response)
6. T: How about, "Why are the snakes called pit vipers?"
7. C: Why do they want to know that they are called pit vipers?
8. T: Try it again.
9. C: Why do they, pit vipers in a pit?
10. T: How about, "Why do they call the snakes pit vipers?"
11. C: Why do they call the snakes pit vipers?
12. T: There you go! Good for you.

Day 4:

TEXT: Spinner's mate is much smaller than she, and his body is dull brown. He spends most of his time setting at one side of her web.

13. C: (No question)
14. T: What's this paragraph about?
15. C: Spinner's mate. How do spinner's mate...
16. T: That's good. Keep going.
17. C: How do spinner's mate is smaller than...How am I going to say that?
18. T: Take your time with it. You want to ask a question about spinner's mate and what he does, beginning with the word "how."
19. C: How do they spend most of his time sitting?
20. T: You're very close. The question would be, "How does spinner's mate spend most of his time?" Now, you ask it.
21. C: How does spinner's mate spend most of his time?

Day 7:

TEXT: Perhaps you are wondering where the lava and other volcanic products come from. Deep within our earth there are pockets of molten rock called *magma.* Forced upward in part by gas pressure, this molten rock continually tries to reach the surface. Eventually—by means of cracks in the crustal rocks or some similar zone of weakness—the magma may break out of the ground. It then flows from the vent as lava, or spews skyward as dense clouds of lava particles.

TABLE 4 *(continued)*

22.	C:	How does the pressure from below push the mass of hot rock against the opening? Is that it?
23.	T:	Not quite. Start your question with, "What happens when?"
24.	C:	What happens when the pressure from below pushes the mass of hot rock against the opening?
25.	T:	Good for you! Good job.

Day 11:

TEXT: One of the most interesting of the insect-eating plans is the Venus's flytrap. This plant lives in only one small area of the world—the coastal marshes of North and South Carolina. The Venus's flytrap doesn't look unusual. Its habits, however, make it truly a plant wonder.

26.	C:	What is the most interesting of the insect eating plants, and where do the plants live at?
27.	T:	Two excellent questions! They are both clear and important questions. Ask us one at a time now.

Day 15:

TEXT: Scientists also come to the South Pole to study the strange lights that glow overhead during the Antarctic night. (It's a cold and lonely world for the few hardy people who "winter over" the polar night.) These "southern lights" are caused by the Earth acting like a magnet on electrical particles in the air. They are clues that may help us understand the Earth's core and the upper edges of its blanket of air.

28.	C:	Why do scientists come to the south pole to study?
29.	T:	Excellent question! That is what this paragraph is all about.

Note: T = teacher, C = student (Charles).

teacher increased her demands accordingly, until the students' behavior became increasingly like that of the adult model, who in turn decreased her level of participation and acted as a supportive audience.

One example of such an interaction is shown in Table 4. This dialogue occurred with a seventh-grade minority student, Charles (IQ = 70, Reading Comprehension grade equivalent = third grade). At the beginning of the training session, Charles was unable to formulate a question. The teacher, estimating that he is having more than usual difficulty with the task, opens her interaction by stating the main idea (Statement 2). She continues to lead him, asking for a "why" question (4) but, receiving no response, she resorts to forming the question for him to mimic (6). Even imitating a fully formed question is difficult for Charles (7, 9). Again, on Day 4, the teacher formulates the question (20), but this time she waits until Charles comes very close to an adequate question by himself. As Charles improves, the teacher demands more from him. On Day 4, the teacher does not open by providing the

TABLE 5
Improvement in Question-Asking by a More Competent Seventh-Grade Student
(Sara)

Day 2:

TEXT: How Can Snakes Be So Flexible?

The snake's skeleton and parts of its body are very flexible—something like a rubber hose with bones. A snake's backbone can have as many as 300 vertebrae, almost ten times as many as a human's. These vertebrae are connected by loose and rubbery tissues that allow easy movement. Because of this bendable, twistable spinal construction, a snake can turn its body in almost any direction at almost any point.

1.   S: Snakes' backbones can have as many as 300 vertebrates—almost _____ times as many as humans.
2.   T: Not a bad beginning, but I would consider that a question about a detail. Try to avoid "fill in the blank" questions. See if next time you can find a main idea question and begin your question with a question word—how, why, when . . . .

Day 3:

TEXT: There are snakes in nearly all parts of the world. Some snakes prefer warm, arid desert areas. Others prefer leafy forests, fields, and woodlands. Some stay in areas near water and are fine swimmers. Then there are several varieties that live all their lives in the sea.

3.   S: Can snakes live their whole lives in seas?
4.   T: See if you can ask a question using your own words.

Day 4:

TEXT: The other kind of camel—the one with two humps—is the *Bactrian.* Its home country is the Gobi Desert of northeastern Asia. The Bactrian has shorter legs and longer wool than the one-humped camel. It also has stronger, more rugged feet. This is important because instead of having sand to walk on, the Bactrian camels live in rough and rocky parts of the world.

5.   S: Where is the Bactrian found?
6.   T: Good for you.

Day 6:

TEXT: When most full-grown spiders want to travel, they have to walk on their eight legs. But some small kinds of spiders, and many young ones, use an easier way. They climb up on bushes, fence posts, or weed stems and spin streamers of silk. When the wind catches the silk and blows it away, each spider tightly holds onto his own streamer. The silk streamer carries him through the air as if it were a parachute or a balloon.

7.   S: I think I have another. When it's traveling, what do they compare the spider to?
8.   T: An interesting question.

TABLE 5 *(continued)*

Day 11:		
TEXT:		The young caterpillar's first meal is its own eggshell. Then it eats a leaf and each day eats more and more food. After a few days, the caterpillar becomes too large for its skin. A new skin forms beneath the first one, the old skin comes open and, like a snake, the caterpillar wriggles its way out of the split skin. Then the caterpillar goes on eating leaves or other kinds of food. When the new skin becomes too tight for the growing body, it again splits and comes off. By then the caterpillar is covered by another skin. This eating and shedding goes on for several weeks. The old skin may be replaced by a new one four or five times. Each time the skin is shed, the size and color of the caterpillar change.
9.	S:	Why does the caterpillar's skin split?
10.	T:	Excellent question. That was the point of the entire paragraph.

Note: T = teacher, S = student (Sara).

main idea, she probes for it (14) and probes for a question (16, 18), which she corrects (20). Note, however, that although the teacher actually produces the questions on both Day 1 and Day 4, on Day 4 she waits until Charles has contributed most of the elements himself.

As Charles' ability to participate increases even further, the teacher again increases the level of participation that she demands from him. On Day 7, she requests a modification to his question form (23), but he formulated the question (24). By Day 11, she receives two excellent questions, but now demands only one (27), that is, she requires him to stick to the exact rules of the game. Finally, by Day 15, Charles can perform his part unaided.

Charles was a particularly weak student at the start, unable to formulate questions at all. Sara, however, began the intervention with a clear notion of the kinds of questions that occur in school — "fill in the blanks." Excerpts from her protocol are shown in Table 5. On Day 2, the teacher, who has tolerated "fill in the blanks" questions until this point, attempts to take the student beyond this level (2) and asks for a main idea rather than a detail question. On Day 3, Sara comes up with a main idea question as requested (3), so again the teacher increases her demands by suggesting that, instead of selecting a line from the text, the student summarize in her own words, a process called invention (Brown & Day, 1983) that is difficult for weaker students to handle. For the remainder of the sessions, Sara's questions are classified primarily as inventions. The teacher has been modeling inventions, and the student has followed suit.

Sara also showed rapid improvement in her generation of summaries. Excerpts from her transcript are shown in Table 6. She begins with detailed bits and pieces of the text (1, 3), progresses to a long description of details (7), and

TABLE 6
Improvement in Summarization by a More Competent Seventh-Grade Student (Sara)

Day 2:

TEXT: How Can Snakes Be So Flexible?

The snake's skeleton and parts of its body are very flexible—something like a rubber hose with bones. A snake's backbone can have as many as 300 vertebrae, almost ten times as many as a human's. These vertebrae are connected by loose and rubbery tissues that allow easy movement. Because of this bendable, twistable spinal construction, a snake can turn its body in almost any direction at almost any point.

1. S: Like, if a snake is turning around, he wouldn't break any bones because he is flexible.
2. T: And the reason he is so flexible is...
3. S: If someone stepped on his tail, he wouldn't be able to move unless he was flexible.
4. T: O.K. That explains why it would be helpful to be so flexible, but what makes the snake so flexible. I think that is an important part of the summary.
5. S: So he can move faster?
6. T: That is another reason for being flexible. Let me do a summary for you. Sara is right. The most important thing we have learned is that the snake is very flexible. The reason they can bend so easily is that they have so many little bones in their backbones and each of them can move separately, making the snake very flexible. Does that make sense?

Day 4:

TEXT: The other kind of camel—the one with two humps—is the *Bactrian*. Its home country is the Gobi Desert of northeastern Asia. The Bactrian has shorter legs and longer wool than the one-humped camel. It also has stronger, more rugged feet. This is important because instead of having sand to walk on, the Bactrian camels live in rough and rocky parts of the world.

7. S: The Bactrian has two humps and it's found in the Gobi Desert of Northeastern Asia and it has short legs and long ... woolier than the other camel and it is stronger and has more rugged feet.
8. T: O.K. And that is because of where it lives. So maybe you could summarize by saying, this is about the second kind of camel which is called the Bactrian camel. It tells us where it lives, how it is built, and why it is built that way.

Day 6:

TEXT: Today salt is plentiful and its use extensive. Science, industry and medicine use salt for a variety of purposes. But it was not always this way.

9.   S: Salt today is plentiful and science and industry use it for a variety of purposes.
10. T: Excellent job.

142

TABLE 6 *(continued)*

Day 7:

TEXT:		Perhaps you are wondering where the lava and other volcanic products come from. Deep within our earth there are pockets of molten rock called *magma*. Forced upward in part by gas pressure, this molten rock continually tries to reach the surface. Eventually—by means of cracks in the crustal rocks or some similar zone of weakness—the magma may break out of the ground. It then flows from the vent as lava, or spews skyward as dense clouds of lava particles.
11.	S:	My summary would be that this paragraph is about magma and magma is molten rock and I wouldn't add anything else.
12.	T:	Good. Good.

Note: T = teacher, S = student (Sara).

by Days 6 and 7 she can summarize an entire paragraph in one sentence (9, 11).

We did not formally score predictions and clarifications because these activities did not occur after every segment, only when the student or teacher thought it appropriate. Students rarely asked for clarification and it should be pointed out that the texts did not contain deliberate ambiguities of the type used in experimental studies of comprehension-monitoring (Baker, 1983; Markman, 1979). Students occasionally asked for the meaning of words. For example, "The word 'prefer' throws me off in this sentence"; "What's the difference between soap and detergent anyway?" "I don't know what 'omitting' is." They sometimes asked for clarification, "I don't see how they can say 'heat lightning occurs on hot summer days.' How could you see it?" "It says here, 'cloud to cloud,' then 'cloud to earth.' Wouldn't that be the same thing?" "I have a question. What do they mean by far away dreams?" And they also discussed their errors, "At first, I didn't get this because I thought the word, 'pumping' was 'bumping,' " and the errors they saw in the texts, "Boy, this paragraph sure is a mess. It's all over the place." In general, they did engage in clarifying and predicting, but not frequently enough to permit formal scoring.

As an objective check on the improvement in dialogue, selections of verbatim transcripts of three sessions for each group were rated by two independent raters. The transcripts, from the beginning, middle, and end of the intervention, were randomized, and the raters' job was to rank them depending on whether they thought the dialogues were from the initial, middle, or final phase of intervention. Percentage of agreement, determined by the number of times the raters correctly identified the order of each transcript, was 83% for the initial and final phases and 67% for the transcripts from the middle segment of training. Given that the students were progressing at different rates, this is reasonable consistency.

Daily Comprehension Assessments: All Groups

The data to be reported here are the percent correct comprehension questions on the daily assessment passages. As there was no reliable effect due to question type, the data were collapsed across this variable. Each day, for 20 instructional days, three of the four poor reader groups (Reciprocal Teaching, Locating Information, and Test Only) answered questions on a novel assessment passage. In addition, the 13 average readers answered questions, each on a randomly selected 5 of the 20 assessment passages.

The average readers scored 75% correct on the passages they read, and this point is included in Figure 1 to provide a normal benchmark against which the poor reading groups can be compared. The poor reading groups' comprehension scores in the major phases of the study (baseline, training first half, training second half, and maintenance) are also shown in Figure 1. It is apparent that the Reciprocal Teaching group improved to the level set by the average readers but that neither the Locating Information intervention (LI) or the Test Only (T) groups improved greatly during the course of the study, and they did not differ from the Control group (C), who received only baseline and maintenance probes. A 3 × 4 mixed analysis of variance was conducted on the scores of the three groups that took the daily assessment (RT, LI, and T), with Groups as a between-subjects variable and Phase (baseline, training first half, training second half, and maintenance) as a within-subjects variable. The main effect of Groups was reliable, $F(2,15) = 4.57, p < .03$, as was the main effect of Phase, $F(3,45) = 13.66, p < .001$. In addition, the Groups × Phase interaction was significant, $F(6,45) = 8.99, p < .001$. This interaction is explained by the fact that, as the intervention progressed, the difference between the reciprocal teaching group and the other conditions grew larger. Confirming our pilot data (Brown & Palincsar, 1982), the reciprocal teaching intervention leads to dramatic improvement in student scores, whereas mere practice taking the tests (T), and even an intervention where the students are helped to answer comprehension questions (LI), do not result in a reliable improvement.[1]

Daily Comprehension Assessments: Reciprocal Teaching

The daily comprehension scores for the Reciprocal Teaching group are shown in Figure 2. Students 1 and 2 received 4 days of baseline, while Stu-

---

[1]In the pilot study (Brown & Palincsar, 1982), locating information did result in a reliable improvement; students improved from 15% to 50% correct and maintained at 40% correct. Our explanation for these differences is that the students in the pilot study were performing so poorly initially that it is doubtful that they understood the task at all, and that the locating information help was sufficient to enable them to improve somewhat. In Study 1, the students began at 40% correct, so they understood the rules of the game, and the locating information was not sufficient to raise their scores significantly.
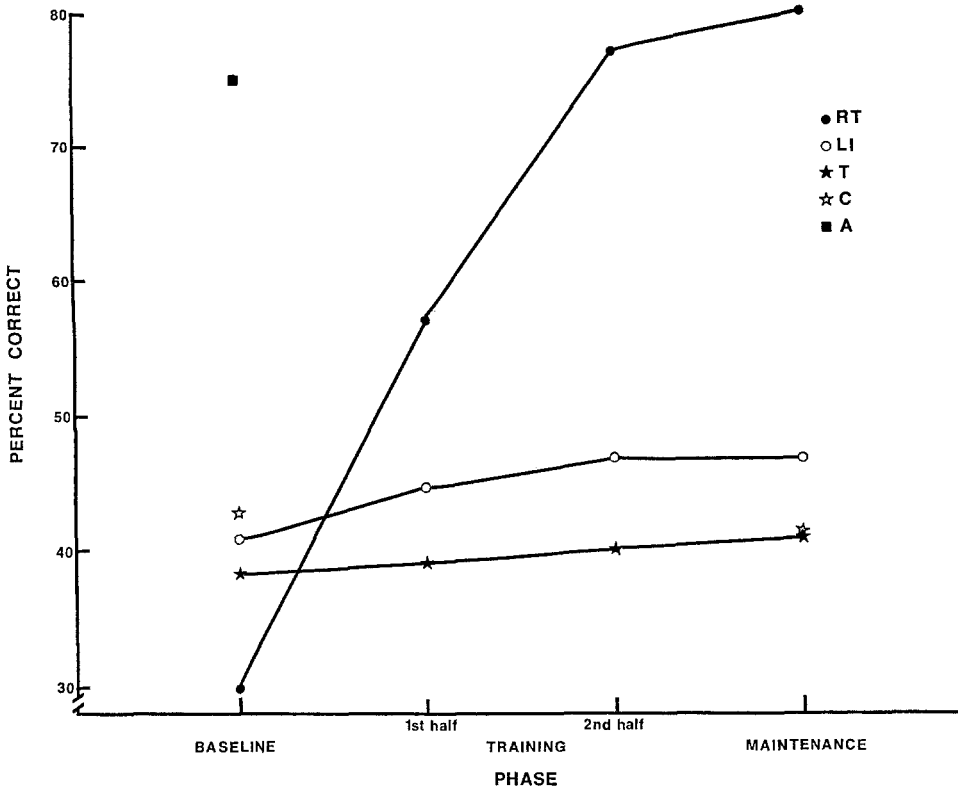
Figure 1.   A comparison of the Reciprocal Teaching (RT) group with the students in the Locating Information (LI), Test Only (T), and untreated Control (C) and the average comprehenders (A) of Study 1.

dents 3 and 4 received 6 days, and Students 5 and 6 received 8 baseline days; in all other respects, the treatment was the same. Visual inspection suggests that the pattern of improvement was similar to that found in the pilot study (Brown & Palincsar, 1982). The six students had baseline accuracy not exceeding 40% correct. Four of the six reached a stable level of 80% correct, taking between 11 and 12 days to do so (Students 1, 3, 4, and 6). Another student, 5, reached a stable level of 70% correct in 12 days. Note that the 13 average readers achieved 75% correct on these passages without intervention. Therefore, for five of the six students, reciprocal teaching experience brought them up to the level set by normal seventh-grade readers. The remaining student, 2, was the only "failure"; however, she did progress from a baseline of 10% and reached a steady level of 50% correct, again in 12 days, a significant improvement—but she never approached the 70 to 80% level of the remaining five students and the normally achieving seventh graders. All of the reciprocal teaching students maintained their improved level of per-
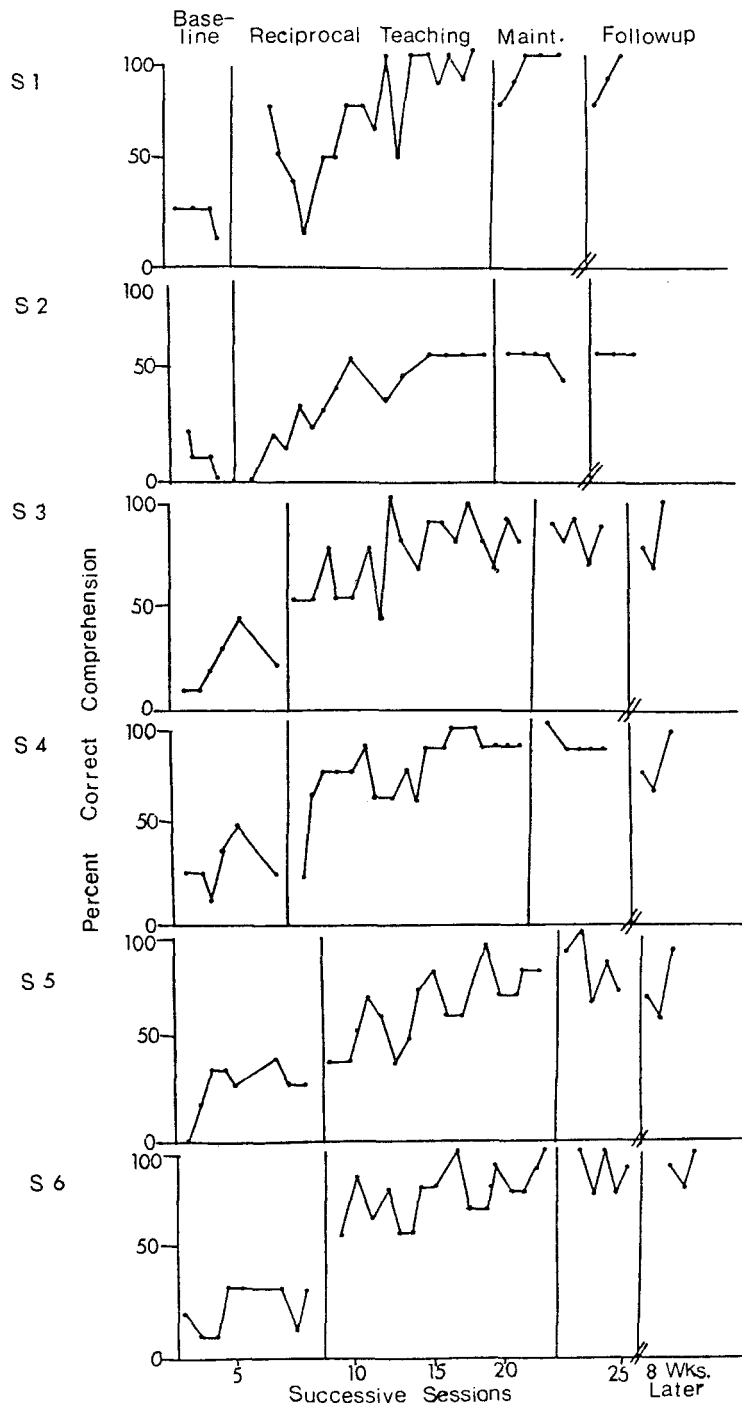
Figure 2. Percent correct on the daily assessment passages for the six subjects in the Reciprocal Teaching group of Study 1.

formance on the maintenance sessions and on the follow-up sessions that took place 8 weeks after the intervention had ceased.

A series of planned comparisons was conducted on the various phases. Confirming the visual impression, mean accuracy during training was significantly higher than during baseline, $F(1,20) = 243, p < .001$. This level was maintained, for there were no significant differences in accuracy between the second half of training and maintenance, or between maintenance and the long-term follow-up that took place 8 weeks after the termination of training.

Generalization probes were taken in the classroom setting for the students in the Reciprocal Teaching group and the untreated Control group. The performance of the reciprocal teaching students was variable, but there was clear evidence of improvement. Probes were taken in two settings, social studies and science. The mean percent correct are show in Figure 3. In general, reciprocal teaching students improved from approximately 20% to 60% correct. This is an impressive finding, particularly given the difficulty investigators have experienced getting generalizable effects of training across task settings (Meichenbaum & Asarnow, 1978; Stokes & Baer, 1977). The only reciprocal
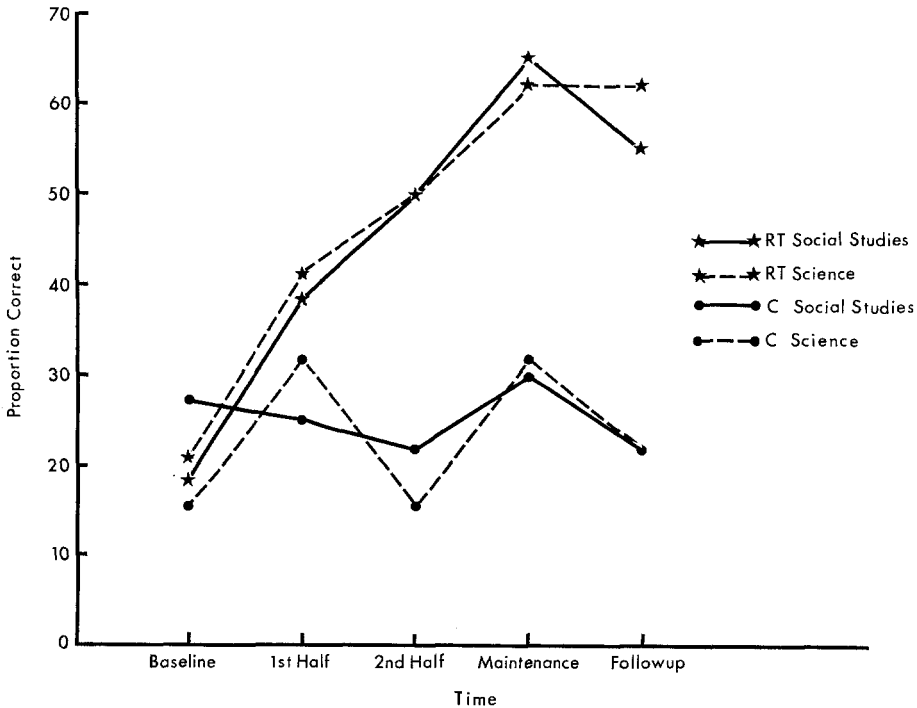

Figure 3.   Percent correct on the classroom generalization probes of Study 1 for the Reciprocal Teaching (RT) and untreated Control (C) groups in both science and social studies classes.

teaching student who did not improve substantially was Student 2, Laura, the same student who reached asymptote at 50% on the daily assessment tests.

Two 2 × 5 mixed analyses of variance were conducted, one on the scores in social studies and the other on the scores in science class. In both, Groups (Reciprocal Teaching vs. Control) was the between-subjects variable, and Phase (baseline, first half, second half, maintenance, and follow-up) was the within-subjects variable. For the social studies data, the main effects of Groups, $F(1,10) = 9.81$, $p < .01$, and Phase, $F(4,40) = 7.27$, $p < .001$, were reliable, as was the Groups × Phase interaction, $F(4,40) = 6.80$, $p < .001$. The Reciprocal Teaching and Control groups started at the same level, but the Reciprocal Teaching group improved over phases while the Control group did not. Similarly, for the science data, the main effects of Groups, $F(1,10) = 18.70$, $p < .001$, and Phase, $F(4,40) = 15.07$, $p < .001$, were reliable, along with the Groups × Phase interaction, $F(4,40) = 7.07$, $p < .001$. Again the groups began at the same level, but the Reciprocal Teaching group showed steady improvement and the Control group did not.

We also include percentile rankings, as this represents a method of showing the students' relative improvement on these classroom tests compared with their nontrained peers. It is a measure of their change in status vis á vis all other seventh graders in the school ($N = 130$). The students' percentile rankings are shown in Table 7. Again, although variable, the performance of the reciprocal teaching students improved, but that of the control students did not. Five of the reciprocal teaching students (Student 2, Laura, again being the exception) demonstrated considerable generalization to the classroom setting, changing their status from the bottom of the distribution to at least average standing.

Transfer Tests

The four transfer tests were administered in pretest-posttest format to two groups, the Reciprocal Teaching group and the untreated Control. In addition, the 13 average readers provided normative data by taking the transfer tests.

Summarizing. The first transfer test to be considered is the summarizing test. Independent raters scored the students' summary sheets and assigned points as follows: One point was given for each trivial or redundant idea that was deleted. One point was awarded for each list of exemplars crossed out, and an additional point for each superordinate given. If a student selected as a topic sentence a unit previously rated as quite important, that was worth one point. If a student selected as a topic sentence a unit previously rated as very important, that was accorded two points. Similarly with inventions, if a student generated an invention rated as quite important, this received one point, and two points were awarded for inventions rated as very

TABLE 7
Percentile Rankings of Students in the Classroom Setting

Class		Social Studies					Science				
Time		Base-line	1st half	2nd half	Main-tenance	Follow-up	Base-line	1st half	2nd half	Main-tenance	Follow-up
Recip-rocal	S1	2	19	39	54	58	2	21	60	62	44
Teaching	S2	2	11	—	11	6	25	21	—	19	20
	S3	27	60	39	80	44	47	26	60	62	82
	S4	43	72	70	69	70	38	53	87	74	82
	S5	1	11	70	69	70	25	21	87	74	37
	S6	5	19	58	80	58	25	53	76	92	44
Median		4	19	58	69	58	25	23	76	68	44
Control	S1	13	11	19	11	6	25	26	32	19	14
	S2	13	47	39	20	17	25	36	32	46	20
	S3	5	31	19	11	6	25	26	8	33	7
	S4	13	6	11	11	17	15	21	8	19	4
	S5	5	6	1	11	2	15	21	32	7	4
	S6	13	11	5	1	6	15	1	8	7	7
Median		13	11	15	11	6	20	24	20	19	7

important. The total number of points earned for the passages summarized during the pretesting and those summarized during posttesting was tallied and used in the analysis. These data are shown in Table 8.

A 2 × 2 mixed analysis of variance, with Groups (RT and C) as the between-subjects variable and Phase (pretest or posttest) as the within-subjects variable, was conducted on the total scores. The Groups × Phase interaction was reliable, F (1,10) = 6.97, $p < .05$. The Control group did not improve, whereas the Reciprocal Teaching group did. The posttest performance of the Reciprocal Teaching group (46.33) was not significantly different from that of the untreated average readers (51.15), $t/y(17)$ = .87.

Perhaps more interesting than the total scores, however, is a profile of the gains of the reciprocal teaching students. Only small gains were made on the superodination rule (6% crossing out lists and 5% naming lists); however, the students were quite facile at this on the pretest. They were not as adept initially at selecting or inventing topic sentences and did improve 20% in these abilities. However, the major gains came in the deletion of redundant and trivial material (33%) and in the importance ratings assigned to their topic sentences (36%). It would appear that the continual instruction during training to paraphrase prose segments by concentrating on the main idea did lead to a significant transfer to a quite dissimilar task.

Predicting questions. The second transfer measure concerned predicting questions. Two independent raters were given typewritten copies of the 10 questions per passage generated by the students (corrected for grammar and spelling). They were asked to rate each question as: a main idea question (worth two points) or a detail question (one point), as a question lifted directly from text (zero points) or paraphrased (one point). In addition, the quality of each question was rated on a five-point scale ranging from one (very poor) to five (excellent). Finally, a question which the rater indicated she would ask herself was awarded an extra point. The raters were trained in the scoring procedure together, and jointly evaluated 10-question sets until they arrived at better than 95% reliability on the rating of the main idea and paraphrase categories. They then each rated all of the remaining questions independently. The Pearson Product Moment correlation coefficient calculated to yield interrater reliability was .88.

The means for the question predicting tasks are also shown in Table 8. A 2 × 2 mixed analysis of variance with Groups (RT and C) as a between-subjects variable and Phases (pretest or posttest) as a within-subjects variable did not result in a reliable interaction effect. However, consider the starting level of these students against the level set by the 13 Control group good comprehenders. The good comprehenders scored 64 points on this task. Prior to training, two of the reciprocal teaching students (Students 3 and 4) also scored above 60, in the normal range, and they did not improve as a

TABLE 8
Performance on Transfer Tests

		Study 1						Study 2	
Group		Average Comprehenders		Control		Reciprocal Teaching		Reciprocal Teaching	
Phase		Pre	Post	Pre	Post	Pre	Post	Pre	Post
Task									
Summarizing	$\overline{X}$	—	51.15	35.50	33.67	36.33	46.33	38.95	48.75
	SD	—	9.19	8.14	9.24	12.01	14.58	10.75	11.70
Predicting Questions	$\overline{X}$	—	63.54	48.75	52.42	44.83	62.63	53.74	61.20
	SD	—	6.29	13.74	10.24	25.03	2.38	7.28	4.51
Detecting Incongruities Total Correct*	$\overline{X}$	—	—	31.00	30.00	29.00	34.33	30.48	33.68
	SD	—	—	4.56	1.55	3.03	1.03	2.78	1.89
Detecting Incongruities Anomalous** Sentences	$\overline{X}$	—	—	1.58	1.35	1.29	1.92	1.26	1.89
	SD	—	—	0.75	0.49	0.56	0.75	0.74	0.57

*Score on Total Correct detection $N = 36$.
**Score on Anomalous Sentences $N/3$ corrected for guessing.

function of training. The remaining four students had lower entering scores (Student 1, 56; Student 2, 31; Student 5, 55; and Student 6, 0), but these four students concluded the study earning an average of 6.17 points. The improvement of the four initially poor students (from 36% to 62%) was reliable ($p < .05$). In addition, the posttest performance of the reciprocal teaching students was not significantly different from that of the average comprehenders, $t (17) = .33$. The poor comprehenders not participating in the reciprocal teaching achieved 48 points on the pretest and 52 on the posttest, not a reliable change.

Detecting incongruities. Several analyses were conducted to compare accuracy detecting congruous and incongruous sentences on the pretest versus the posttest. There were a total of 36 lines evaluated by each student in each phase; 3 lines were in fact incongruous. Consider first the total ($N = 36$) correctly identified. These data are shown in Table 8. A 2 × 2 mixed analysis of variance, with Groups (RT and C) as a between-subjects variable and Phase (pretest and posttest) as a within-subjects variable, was conducted on the total number of lines corrrectly identified. The Groups × Phase interaction was reliable, $F (1,10) .eq 10.44, p < .01$. The Control group did not improve (31–30), but the Reciprocal Teaching Group did (29–34).

A second measure of accuracy is the number of correct detections of incongruous sentences. Of the 36 lines, 3 were objectively anomalous. Accuracy on this measure was defined as the number of correct detections of incongruous sentences minus three times the percentage of times a student said "no" when evaluating if the sentence made sense in the story, a correction for guessing. Corrected for guessing, the mean pretest score for the reciprocal teaching students was 1.29 correct, and the mean posttest score was 1.92 correct recognition of incongruous sentences. This improvement was significant, $t (5) = 4.79, p < .001$. The Control group, however, did not improve (1.58 to 1.35).

Not only did the rate of detecting incongruities improve in the Reciprocal Teaching group, but so also did the quality of verbal responses during probes. When the students indicated a line didn't make sense in the story, the investigator would probe, "Can you tell me why this line doesn't make sense in the story?" During pretesting, very typical responses to the probe were, "It just doesn't read right," "It isn't important," "They need to be more specific," "I just don't like that one," or "I don't know." In contrast, on the posttest, even when the students incorrectly evaluated a line as not making sense, their reasons were generally more specific and informative, for example, "They said the boat was in the water, so how can it be off the shore?" "The recipe didn't tell you what they were making," or "Under a slide they usually have sand, not grass." One comment which teachers would appreciate was made

several times when students were appraising *Day in the Park*: "It says, 'the teacher is very tired,' but teachers don't get tired," or more enlightening still, "It says, 'the teacher is very tired,' but they don't do anything."

Rating importance.   The data for this transfer test will not be reported in detail. The test was simply a failure. The students showed no preference for any level of importance, distributing their rankings randomly, both on pretest and posttest. This performance was in sharp contrast to that of the 13 Control group average comprehenders who, replicating the original Brown and Smiley (1977) seventh-grade data, rated items in concordance with college students. Note that the students in this study did improve in their ability to select important topic sentences in the summarizing task. Thus, we assume that the difficult Brown and Smiley rating task was an insensitive measure for poor students of the emergent ability to concentrate on importance at the expense of trivia, demanding as it does that students simultaneously keep in mind large segments of text and rate each and every one of the idea units in terms of fine degrees of importance.

Standardized Scores

After the completion of the study, approximately 3 months after the termination of intervention, we were able to administer the Gates-MacGinitie Standardized Reading Test to the reciprocal teaching students. This readministration was conducted on a one-to-one basis by an investigator, whereas the students' existing standardized scores were group-administered by the school. There was no improvement on the vocabulary measure (mean gain = 1 month). On the comprehension measure, however, four of the six reciprocal teaching students made substantial gains. These gains are shown in Table 9. The average gain, including all six students, was 15 months.

Case Studies: The Relation of Qualitative and Quantitative Improvement Measures

If we consider the six students in the Reciprocal Teaching group as individual cases, then a close relation between improvement in the dialogues and improvement in the assessment passages and transfer tests can be seen. Although they differed in terms of starting competence, and responded to the instruction at different rates, in all cases their individual progress began with a gradual adoption and refinement in the dialogues of the procedures modeled by the teacher. Improvement was first shown in the dialogues after which it was reflected in the students' individual scores, a pattern of results that suggests a gradual internalization (Vygotsky, 1978) of behavior originally perfected in a social setting. The reciprocal teaching dialogue provided

TABLE 9
Improvement in Standardized Comprehension Scores of the
Reciprocal Teaching Students

Grade Equivalent	Pre-Intervention Grade + Months	Post-Intervention Grade + Months	Gain (in months)
Subject			
1	4,2	4,4	+2
2	3,6	3,6	+0
3	5,3	8,3	+36
4	6,4	7,9	+17
5	3,9	5,0	+15
6	3,1	4,9	+20

the contextual support for the first emergence of the skill, students and teachers providing examples, support, and feedback for each other. After a new skill was practiced in the group setting, the student then became able to apply it independently on the daily assessments and classroom generalization tests. We will illustrate this with some examples.

Charles (Student 6) was clearly a success story. A minority student, he had a low IQ (70), and was a full 4 years delayed on standardized reading comprehension scores at the outset of the study. He scored 30% correct during the baseline assessments and 20% in both of the initial classroom probes. He earned no points on the question generating pretest as he persisted in making statements rather than posing questions. Of the six students, he earned the fewest number of points for his summaries written during pretesting.

In the reciprocal teaching sessions, he began slowly (see Table 4), unable to form a question until the ninth time he was called upon to do so (Day 2). However, Charles' ability to generate questions continued to improve on each day of reciprocal teaching, as did his summarization skills. By Day 5, Charles was achieving a 50% success rate with his dialogues; and by Day 10, he had improved his success rate to 75%, a level he continued to maintain.

The link between improvement in the dialogues and the independent assessments of Charles' skill is clear. He proceeded from a starting level of 30% correct on the assessment passages to 60% correct by Day 5, the day on which he was able to generate adequate summaries and questions on 50% of the occasions he was called upon to do so. He reached asymptotic level in the dialogues (75% of questions and summaries judged perfectly adequate) on Day 10, and by Day 15 he no longer required any help. Similarly, by Day 10, Charles reached 70% and by Day 15, 80% to 90% correct (asymptote) on the assessment passages.

The improvement shown in the daily dialogues and assessment passages was also seen in the posttests. For example, in addition to the stable 80% to 90% correct on the daily passages, Charles improved from 20% to 60% correct on the classroom generalization probes, and from 0 to 65 points on the question-generating transfer test, and 20 months on the Gates-MacGinitie Standardized Comprehension Score.

Alice (Student 1) was also a poor student at the outset. The difficulty she experienced during laboratory baseline measures (30%) was reflected in the classroom as well (5%). Her science teacher, in fact, commented that "she is a good kid, but . . ." He noted that all she did was sit there during class, and he was doubtful that much could be done to help her. However, Alice, who was delighted to participate in the study, gleefully reported after several days of intervention that she had been the only one in history class who had been able to define what a summary was. "I raised my hand, the teacher looked around, but I was the only one. He had to call on me, and I got it right!"

Alice showed almost immediate improvement on both her dialogues and her daily assessments. By Day 8, she was achieving a 70% success rate on her questions and summaries in the dialogues, and this was the day on which she began scoring consistently at 70% correct on the daily tests. By Day 12, 87% of her dialogue summaries and questions were approved by the teacher, and then she began her stable rate of responding between 80% and 90% correct on the daily passages. And on classroom generalization probes, she also improved from 5% to 60% correct, moving her from the 2nd to 50th percentile for her class. Her transfer performance was also comparable to normal achieving seventh graders. However, she failed to show improvement on the standardized tests of comprehension.

Mary's (Student 5) level of starting competence was different from that of Charles and Alice. Although her comprehension scores on the assessment baselines were very low, both in the laboratory (0%) and in the classroom probe (10%), she did show considerable ability to handle the "transfer" pretests of summarization and questioning. Of the six reciprocal teaching students, Mary earned the greatest number of points for her pretest summaries (53) and earned a fair number of points for her questions as well (55). This ability to formulate questions and summaries was also seen in the dialogues. Even in the first five days, 90% of her questions and 95% of her summaries were judged adequate. Mary's challenge then was somewhat different from that of Charles and Alice. She needed to learn how to apply her summarizing and questioning skills to the task of fostering comprehension while reading independently, that is, improve her comprehension test performance in independent reading sessions. And she did. By the second half of training, Mary was scoring 70% to 80% on the daily assessments, and 60% and above in both social studies and science, placing her above the 70th percentile in both classes. She improved her standardized comprehension score by 15 months.

Sara (Student 3) also had little difficulty engaging in the dialogues; by Day 3 all of her responses were judged to concern the gist, and by Day 7 they were judged to be perfectly adequate. On the daily assessments, she progressed from 20% to 70% correct by Day 7, and she maintained this level until Day 11, when she reached asymptote of 80% to 90% correct. This progress on independent tests was also reflected in the classroom but with less consistency and some delay, for it was not until the last half of training that she scored well above her baseline in both science and social studies, placing her in the 80th and 62nd percentiles, respectively. Sara also increased her standardized comprehension score by 36 months, bringing her to 15 months *above* grade level.

Even Laura (Student 2), the only student who failed to reach the level (75%) set by the normal readers, showed evidence of a gradual improvement in the group dialogues that was reflected in her independent work. Although her questions remained poor, her summaries improved from 32% to 75% judged as adequate. This higher level was reached by the middle of the sessions, and on Day 12 Laura reached her asymptotic level on the daily assessments, 50% correct, a significant improvement over her starting level of 10% correct.

Although the students had different problems and progressed at different rates, the pattern of improvement was similar. First an improvement was shown in the dialogues and then the independent scores began to improve. Skills practiced in the social context came to be used independently by the students.

Summary of Study 1.   The main findings of Study 1 are that students responded very well to the reciprocal teaching intervention, gradually performing more and more like the adult model and becoming better able to take their turn as dialogue leader. Main idea questions and summaries came to predominate, inventions in their own words replaced verbatim selections from the text, and incomplete, unclear, and detailed responses dropped out. The teacher guided this progress by demanding that students perform slightly above the level they had achieved previously, thus providing them with a reasonable challenge, a chance to take an additional step in their zones of proximal development (Vygotsky, 1978).

In addition to the changes in their participation in the reciprocal teaching sessions, the students improved dramatically on their daily assessment passages. All students reached asymptote within 12 days and, for five of the six, the level was at 70% to 80% correct, comparable to accuracy attained by the average comprehenders who acted as control subjects. Only Student 2 failed to reach the normal level, but she did improve from 10% to 50% and maintained that level; indeed, all of the students maintained their asymptotic level for at least eight weeks.

In addition to these increases on the daily comprehension measures, five of the six students improved their classroom comprehension scores from 20% to 60% correct, resulting in a substantial change in their relative class status, as reflected in their percentile rankings.

There was also encouraging evidence of transfer to new tasks. Reliable improvement was found in the ability to use condensation rules for summarizing, in the ability to predict questions that a teacher might ask concerning a text segment, and in the ability to detect incongruous sentences embedded in prose passages.

STUDY 2

Given the success of Study 1 and the original pilot investigation (Brown & Palincsar, 1982), we decided to attempt another replication, but this time the teacher would be a "real" teacher, not an investigator, and the instruction would take place in naturally occurring groups within school settings. The practical significance of any intervention is dramatically improved if it can be implemented under conditions approaching those of the normal classroom. In Study 2, we looked at four groups of students, two classroom reading groups for the poorest readers, and two remedial reading groups that met regularly in a resource room. In all other respects, the study was a replica of Study 1.

Method

Subjects. The students were seen in the reading groups they normally attended. Originally, six teachers were contacted, and five were willing to participate. After screening the students assigned to these teachers, one group was dropped because the students did not meet the decoding requirements for entry into training. Of the remaining four groups, two were taught by regular classroom teachers (Groups 1 and 4) and two by resource room teachers (Groups 2 and 3), all in rural schools in central Illinois. Groups 2 and 3 consisted of five students each, Group 1 of seven students and Group 4 of four students. Group 1 received 4 days; Group 2, 6 days; Group 3, 8 days; and Group 4, 10 days of baseline. In all other respects, they were treated identically.

The majority of the students were seventh graders. However, Group 4 was an eighth-grade grouping, one student in Group 1 was an eighth grader, and two students of Group 2 were sixth graders. Seven of the 21 students were female, all were white. The standardized scores for each subject are shown in Table 10. All students met the decoding criteria of 80 wpm correct with no more than two error words per minute when orally reading seventh-grade

texts. All students scored 50% or less correct on the baseline assessment com-
prehension passages. Their standardized comprehension scores were varia-
ble, ranging from approximately at grade level to several years delayed. In
general, however, the students averaged two and a half years delayed in read-
ing comprehension scores.

   To summarize, all the subjects were junior high students identified by their
schools as requiring supplementary or special reading instruction. They were
not labeled as learning disabled or mentally retarded. Standardized testing
and our own inventory indicated that they all met the minimum decoding
standard to permit instruction on seventh-grade texts (Lovitt & Hansen,
1976a), but they were poor comprehenders, typically comprehending at
about 2 years below grade level. The students were all instructed by experi-
enced teachers in their natural groups. They shared educational and environ-
mental backgrounds more alike than dissimilar.

Procedure

With the exception that the training was conducted by the teachers, the pro-
cedures and materials for Study 2 were identical to those of Study 1. The
teachers themselves received three training sessions. In the first, they were in-
troduced to the rationale behind the reciprocal teaching intervention and
shown the results of the pilot study. They also viewed and discussed a
videotape of the first author employing the technique with a group of
students.

   In the second training session, the first author practiced the procedures
with each teacher. The first author modeled both the teacher's role and be-
haviors that might be expected from students. Difficulties that could arise
were anticipated and discussed, such as situations where a student is unable
to generate a question, or where a student summarizes by reiterating the
whole paragraph in detail. Remedial steps were demonstrated.

   In the final session, each teacher met with the first author and a group of
seventh graders who were not taking part in the study, and they practiced the
procedure together. The first author modeled how the procedure should be
introduced to the students, modeled the four main dialogue activities, and
the process of feedback. The teacher then assumed responsibility for the
group and, as the practice session transpired, the teacher and investigator dis-
cussed the proceedings with one another. In addition, the teachers were left
with several pages of directions regarding the introduction and daily format
of the training sessions. The first author also checked weekly on the teacher-
directed sessions to see if the intervention was being conducted properly.
These visits provided further opportunity for discussion and resolution of
any difficulties encounted. All reciprocal teaching sessions were tape-
recorded.

<table>
<thead>
<tr><th>Subject #</th><th colspan="2">Comprehension</th><th colspan="2">Vocabulary</th><th colspan="2">Oral Reading, wpm</th><th>Baseline</th></tr>
<tr><th></th><th>GE[1]</th><th>Delay</th><th>GE[1]</th><th>Delay</th><th>Correct</th><th>Incorrect</th><th>% Correct</th></tr>
</thead>
<tbody>
<tr><td>Group 1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>1</td><td>6,0</td><td>−1,0</td><td>5,4</td><td>−1,8</td><td>80</td><td>0.8</td><td>20</td></tr>
<tr><td>2</td><td>3,9</td><td>−3,3</td><td>4,9</td><td>−2,3</td><td>85</td><td>1.5</td><td>30</td></tr>
<tr><td>3</td><td>4,2</td><td>−2,10</td><td>5,4</td><td>−1,8</td><td>81</td><td>1.8</td><td>50</td></tr>
<tr><td>4</td><td>*5,7</td><td>−2,5</td><td>7,2</td><td>−0,10</td><td>87</td><td>0.8</td><td>50</td></tr>
<tr><td>5</td><td>6,2</td><td>−0,10</td><td>6,2</td><td>−0,10</td><td>98</td><td>1.2</td><td>50</td></tr>
<tr><td>6</td><td>2,0</td><td>−5,0</td><td>4,9</td><td>−2,3</td><td>84</td><td>1.9</td><td>50</td></tr>
<tr><td>7</td><td>4,0</td><td>−3,0</td><td>6,5</td><td>−0,7</td><td>97</td><td>0.9</td><td>50</td></tr>
<tr><td>X̄</td><td>4,6</td><td>−2,7</td><td>5,8</td><td>−1,5</td><td>87</td><td>1.3</td><td>43</td></tr>
<tr><td>Group 2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>8</td><td>5,2</td><td>−1,10</td><td>6,1</td><td>−0,11</td><td>113</td><td>2.0</td><td>30</td></tr>
<tr><td>9</td><td>*3,8</td><td>−2,4</td><td>5,4</td><td>−0,8</td><td>145</td><td>0.8</td><td>30</td></tr>
<tr><td>10</td><td>4,5</td><td>−2,7</td><td>3,6</td><td>−3,6</td><td>80</td><td>1.8</td><td>20</td></tr>
<tr><td>11</td><td>*5,9</td><td>−0,3</td><td>4,5</td><td>−1,7</td><td>109</td><td>2.0</td><td>30</td></tr>
<tr><td>12</td><td>3,8</td><td>−3,4</td><td>5,8</td><td>−1,4</td><td>81</td><td>0.9</td><td>20</td></tr>
<tr><td>X̄</td><td>4,6</td><td>−2,1</td><td>5,0</td><td>−1,7</td><td>106</td><td>1.5</td><td>26</td></tr>
<tr><td>Group 3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>13</td><td>4,6</td><td>−2,6</td><td>5,1</td><td>−1,11</td><td>113</td><td>2.0</td><td>30</td></tr>
<tr><td>14</td><td>4,6</td><td>−2,6</td><td>5,9</td><td>−1,3</td><td>82</td><td>1.4</td><td>30</td></tr>
<tr><td>15</td><td>5,3</td><td>−1,9</td><td>6,5</td><td>−0,7</td><td>100</td><td>1.1</td><td>20</td></tr>
<tr><td>16</td><td>4,7</td><td>−2,5</td><td>7,1</td><td>+0,1</td><td>88</td><td>2.0</td><td>30</td></tr>
<tr><td>17</td><td>6,0</td><td>−1,0</td><td>4,9</td><td>−2,3</td><td>136</td><td>1.6</td><td>20</td></tr>
<tr><td>X̄</td><td>5,0</td><td>−2,0</td><td>5,10</td><td>−1,2</td><td>104</td><td>1.6</td><td>26</td></tr>
<tr><td>Group 4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>18</td><td>*4,7</td><td>−3,5</td><td>7,0</td><td>−1,0</td><td>113</td><td>2.0</td><td>30</td></tr>
<tr><td>19</td><td>*6,6</td><td>−1,6</td><td>7,5</td><td>−0,7</td><td>129</td><td>0.7</td><td>30</td></tr>
<tr><td>20</td><td>*7,4</td><td>−0,8</td><td>7,6</td><td>−0,6</td><td>119</td><td>2.0</td><td>20</td></tr>
<tr><td>21</td><td>*3,9</td><td>−4,3</td><td>6,5</td><td>−1,7</td><td>129</td><td>0.7</td><td>30</td></tr>
<tr><td>X̄</td><td>5,6</td><td>−2,6</td><td>7,1</td><td>−0,11</td><td>122</td><td>1.4</td><td>28</td></tr>
</tbody>
</table>

*All subjects except those marked with * are seventh graders. Subject 4 of Group 1 was an eighth grader, as were all the subjects of Group 4. Subjects 9 and 11 of Group 2 were sixth graders.

[1]Grade Equivalent, Gates-MacGinitie Test. Grade Equivalent Score = grade plus additional months, i.e., 4,2 = fourth grade plus 2 months.

The students were shown their progress charts on a daily basis during baseline, maintenance, and long-term follow-up, and on a weekly basis during intervention, and their improvement was discussed with them.

RESULTS AND DISCUSSION

Quality of dialogue.   A similar improvement in quality of dialogue was found as in the pilot study (Brown & Palincsar, 1982) and Study 1, but was somewhat less dramatic in Study 2. In the group settings, the teachers called upon the "better students" in the initial sessions and then gradually introduced the poorer students into the dialogue as they felt they could handle the responsibility, a natural procedure for experienced teachers. This resulted in a group level of student response that was higher initially and did not improve as dramatically over sessions as did the individuals of Study 1, the group effect to some extent masking individual learning. Even so, the same trend was observed with incomplete or unclear questions descreasing significantly from 20% to 4% ($p < .001$) and main idea questions increasing (though not significantly) from 57% to 70% across the sessions. Similarly, main idea summaries increased from 68% to 85% of the total produced by the groups ($p < .01$).

Because of the larger group size, the students were able to provide modeling and feedback for each other, learning from their peers as well as from their teacher. As the sessions progressed, the teacher was able to hand over a great deal of the work to the students and serve as a coach. An illustration of this change is shown in the contrast between the dialogue samples in Table 11, taken from an early session (Day 3), where the teacher is in control and the students interact almost exclusively with her, and that of a later session, Day 13, from the same group, shown in Table 12, where the students respond to each other with encouragement from the adult teacher.

Both the Day 3 and Day 13 dialogues attest to the fact that the students ($N = 5$) and teacher were able to engage in a smooth flowing discussion. On Day 3, however, the teacher is very much the pivotal participant. As can be seen in Table 11, one session of the silent reading is followed by one extensive dialogue, where the students interact with one another only once (statements 1 to 3); the remainder of the runs are S-T, S-T, student followed by teacher. In the table, the teacher contributions are marked with an asterisk to highlight this pattern. The students interact individually with the teacher, not with each other. Note also that the entire interaction focuses on one segment of text and on one disputed point — the use of snakes' tongues. Interestingly, another group had problems with this segment, one student reading, "no snake's tongue is completely harmless," instead of the correct, "No, — snakes' tongues are completely harmless," thus generating an interesting confusion and occasion for clarification.

TABLE 11
Reciprocal Dialogue from an Early Session; Study 2, Group 3, Day 3

Text:		Can Snakes Sting With Their Tongues?
		No—snakes' tongues are completely harmless. They're used for feeling things and for sharpening the snakes' sense of smell. Although snakes can smell in the usual way, the tongue flickering in the air picks up tiny particles of matter. These particles are deposited in two tiny cavities at the base of the nostrils to increase the snakes's ability to smell.
1.	A:	Do snakes' tongues sting?
2.	K:	Sometimes.
3.	A:	Correct. This paragraph is about do snakes sting with their tongue, and different ways that the tongue is for and the senses of smell.
4.	*T:	Are there any questions?
5.	C:	Snakes' tongues don't sting.
6.	*T:	Beautiful! I thought, boy, I must have been doing some fast reading there because I missed that point. A _____, could you ask your question again?
7.	A:	Do snakes' tongues really sting?
8.	*T:	Now, A _____, since you have asked the question, can you find in that paragraph where the question is answered?
9.	A:	No, snakes' tongues are completely harmless.
10.	*T:	So we'll try it again. Can you generate another question that you think a teacher might ask?
11.	A:	What are the tongues used for?
12.	*T:	Good!
13.	L:	The sense of smell.
14.	*T:	Is that correct? A _____, do you disagree? Yes.
15.	A:	That answer was right, but there are other things that the tongue can do.
16.	L:	But she only said tell one, she didn't say tell all of them.
17.	*T:	O.K.
18.	B:	It is used to pick up tiny particles.
19.	*T:	O.K. I think that this is an important point. You have the basic concept which is correct, O.K., but what the question really is saying is, is it used for smell? O.K.?
20.	B:	They are used for feeling things for sharpening snakes' sense of small.
21.	*T:	O.K. They are used for sharpening the snakes' sense of smell. Are they used for smelling? That's the point we aren't clear on.
22.	L:	In my answer I said it is for the sense of smell.
23.	*T:	That is fine; this is what the technique is all about. What it means is not that you are right or wrong or good or bad. What it says is that we have just read something and have had a disagreement about what it says. We need to work it out.
24.	A:	My prediction is that they will now talk about the different things about snakes. Where they live, and what they eat and stuff like that.
25.	*T:	O.K. Good. What do you think they eat?
26.	A:	Animals.
27.	A:	Deserts.
28.	C:	I challenge. Snakes, all of them, don't live in the desert. They live in the woods and in swamp areas.

Note: *indicates turns taken by adult teacher (T); A, B, C, K, and L are students.

Text:		In the United States salt is produced by three basic methods: solar (sun) evaporation, mining, and artificial heat evaporation. For salt to be extracted by solar evaporation, the weather must be hot and dry. Thus, solar salt is harvested in the tropic-like areas along our southern ocean coasts and at Great Salt Lake.
1.	C:	Name three different basic methods how salt is produced.
2.	A:	Evaporation, mining, evaporation ... artificial heat evaporation.
3.	C:	Correct, very good. My summary on this paragraph is about ways that salt is being produced.
4.	*T:	Very good. Could you select the next teacher? (Student selects another student)
Text:		The second oldest form of salt production is mining. Unlike early methods that made the work extremely dangerous and difficult, today's methods use special machinery, and salt mining is easier and safer. The old expression "back to the salt mine" no longer applies.
5.	L:	Name two words that often describe mining salt in the old days.
6.	K:	Back to the salt mines?
7.	L:	No. Angela?
8.	A:	Dangerous and difficult.
9.	L:	Correct. This paragraph is all about comparing the old mining of salt and today's mining of salt.
10.	*T:	Beautiful!
11.	L:	I have a prediction to make.
12.	*T:	Good.
13.	L:	I think it might tell when salt was first discovered, well, it might tell what salt is made of and how it's made.
14.	*T:	O.K. Can we have another teacher?
Text:		Table salt is made by the third method—artificial evaporation. Pumping water into an underground salt bed dissolves the salt to make a brine that is brought to the surface. After purification at high temperatures, the salt is ready for our tables.
15.	K:	After purification at high temperatures the salt is ready for what?
16.	C:	Our tables.
17.	K:	That's correct. To summarize: After its purification, the salt is put on our tables.
18.	*T:	That was a fine job, Ken, and I appreciate all that work, but I think there might be something else to add to our summary. There is more important information that I think we need to include. This paragraph is mostly about what?
19.	A:	The third method of artificial evaporation.
20.	B:	It mainly tells about pumping water from an underground salt bed that dissolves the salt to make a brine that is brought to the surface.
21.	*T:	Angela hit it right on the money. This paragraph is mostly about the method of artificial evaporation and then everything else in the paragraph is telling us about that process. O.K. Next teacher.

Text:		For thousands of years people have known salt—tasting it, using it for their lives' needs. Sometimes it has been treasured as gold; other times it has been superstitiously tossed over the shoulder to ward off bad luck. Somehow people and salt have always been together, but never is the tie more complete than when the best people are called "the salt of the earth."
22.	C:	My question is, what are the best people called?
23.	L:	The salt of the earth.
24.	C:	Why?
25.	L:	Because salt and the people have been together so long.
26.	*T:	Chris, do you have something to add to that? O.K. It really isn't because they have been together so long; it has to do with something else. Brian?
27.	B:	(reading) "People and salt have always been together but never has the tie been so complete."
28.	*T:	Alright, but when we use the expression, "That person is the salt of the earth," we know that means that person is a good person. How do we know that?
29.	B:	Because we treasure salt, like gold.

Note: *indicates turns taken by adult teacher (T); A, B, C, K, and L are students.

The same group is seen again, 10 intervention days later, in the dialogue shown in Table 12. Here, four reading-dialogue sets are included in 29 statements, rather than only one as in Table 11. Now the majority of the "runs" are student-controlled, with the teacher interspersing praise and encouragement (4, 10, 12) and some management (4, 14, 21). The teacher only intercedes with advice and modeling when a student misses the point and the other students do not catch it (statements 18, 26, 28). The teacher has moved from the pivotal role of responding individually to each child, to a coach who sits in the background, offers encouragement, and occasionally pushes for a better interpretation of the text. The expert provides just the degree of scaffolding necessary for the dialogues to remain on track, leaving the students to take as much responsibility as they can.

Daily comprehension assessment. Again, no differences due to question type (text-explicit, etc.) were found and, therefore, the data were collapsed across this variable. The four groups of students were subjected to different amounts of baseline (from 4 to 10 days), otherwise they were treated identically.

Individually, the students performed in a manner similar to that found in Study 1. All of the students in Groups 1 through 3 individually reached criterion within 15 days. In Group 4, all students reached criterion in 5 days. If one considers the group means, two groups reached criterion in 13 days (Groups 1 and 2), one in 9 days (Group 3), and one in 5 days (Group 4). It is

interesting to note that, in Group 4 (the only eighth-grade grouping), two of the four students were performing excellently on the first day. The resultant group, in some sense, consisted of three models, the teacher and the two good students, and two tutees, the remaining two poor students. In this favorable milieu, the poor students rapidly improved, and the entire group reached criterion in 5 days, versus a mean of 12 days for the other groups and the individual students of Study 1. Such findings, if replicated, could have important implications for decisions concerning the composition of the "optimal" reading group.

The daily comprehension means per group are shown in Figure 4. Students were typically achieving 40% accuracy on comprehension questions during baseline. With the introduction of the intervention, their accuracy increased steadily, until all students were consistently scoring between 70% and 80% by the 15th day of intervention. The students continued to show gains during maintenance, with slight decrements during follow-up.

To confirm these observations, phase contrasts were conducted. Mean accuracy during training was significantly better than during baseline, $F(1,80) = 487, p < .001$, as was accuracy in the second versus the first half of training, $F(1.80) = 76.701, p < .001$. Unlike Study 1, performance continued to increase during maintenance, $F(1,80) = 5.72, p < .02$. Although this difference is reliable, it represents only a 3 percentage point increase. There was a slight decline at the eight-week follow-up compared with immediate maintenance performance, $F(1,80) = 7.61, p < .01$, but again, this represented a five point difference in actual scores. Performance on follow-up was equivalent to the last few days of training. In short, students in Study 2 started at a level of approximately 40% correct and ended at a level of 80% correct, an impressive effect of training.

Transfer tests.   There were no differences between groups in their performance on the transfer tests; therefore, the data are collapsed across this variable. The same pattern of transfer results occurred in Study 2 as for the reciprocal teaching group of Study 1. The rating thematic importance task again resulted in random performance both before and after intervention, and will not be discussed further here (see Palincsar, 1982, for full details). The remaining three transfer tests resulted in reliable improvements. The data are shown in Table 8, along with that of the average students and the Reciprocal Teaching group of Study 1.

Summarizing.   The difference between pretest and posttest scores (38.95 and 48.71, respectively) was significant, $t(20) = 3.24, p < .004$. The largest gain again occurred in the rating importance of topic sentences (40%). These students also improved (19%) on the most difficult rule, inventing a topic sentence. For example, after reading a seven sentence paragraph about
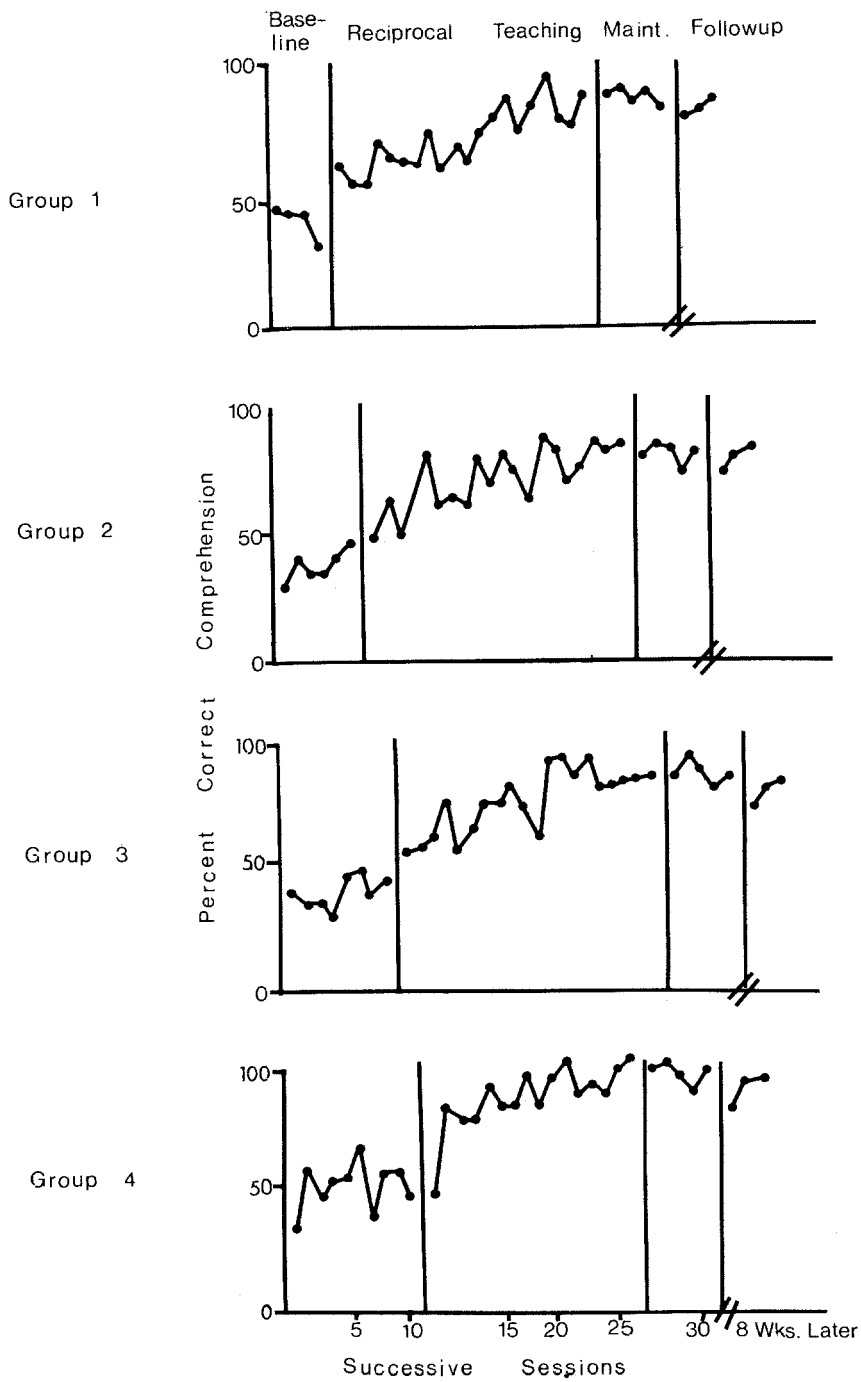
Figure 4. Group mean percent correct on the daily assessment passages for the class-room groups of Study 2.

two different men who kept tarantulas (one to deal with cockroaches in the kitchen and another to keep robbers away from his jewelry store), one student crossed out the paragraph and wrote, "Spiders can get rid of disturbers." The tarantulas were not referred to as spiders in the passage, and the words "disturbers" or "disturb" did not appear in the text.

Predicting questions.   Out of the total possible score of 100, the students scored a mean of 53.62 on pretest and 61.24 on posttest, a significant difference, t (20) = 4.58, $p < .001$. The percentage of gain was distributed in the following manner. The greatest percentage of gain points was earned for quality of questions (62%). Raters determined that posttest questions were more clear and complete than pretest questions. Twenty-one percent of the gain was accounted for by the increased likelihood of the students asking questions similar to those proposed by the raters. Thirteen percent of the gain was attributed to asking more main idea than detail questions. The posttest level of 61.24 points is quite similar to the posttest level of 62.50 from Study 1 and the 64.0 level of the 13 average comprehenders included as comparison students in Study 1. Again, the improvement in posttest scores after training is modest but reliable, bringing the poor comprehenders up to the "normal" level.

Detecting incongruities.   The mean numbers of lines per passage appropriately classified ($N = 36$) or recognized as anomalous ($N/3$) are also shown in Table 8. The total identification difference between the pretest (30.47) and the posttest (33.47) was reliable, t (20) = 7.44, $p < .001$. The mean number of incongruous sentences detected (corrected for guessing) rose from 1.26 on the pretest to 1.84 on the posttest, again a reliable difference, t (20) = 5.60, $p < .001$, similar to that found in Study 1.

Summary of Study 2.   Very similar results were found in Studies 1 and 2. The effect of the reciprocal teaching intervention was reliable, durable, and transferred to tasks other than the training vehicle. The similarity of the main results across the experiments is more striking than the differences. Even though the intervention was one-to-one in the pilot study (Brown & Palincsar, 1982), in student dyads in Study 1, and in larger, naturally occurring, groups in Study 2, the same pattern of results obtain. Classroom teachers, receiving only limited introduction to the method, were as effective as the first author in conducting the intervention.

It is interesting to note that, without exception, the teachers expressed a degree of skepticism regarding their students' ability to participate in the reciprocal teaching prior to beginning the study. At the conclusion of the study, the teachers were pleased not only with the progress demonstrated by the students in the reciprocal activities and their improvement with the comprehen-

sion measures, but by other results as well. The teachers observed that general "thinking" skills seemed to improve. The students appeared better able to locate important information and to organize their ideas — skills which the teachers regarded as important "study skills."

The students also reported that they were using the instructed activities (primarily summarizing and question predicting) in their content classes. As one student proudly reported to his reading teacher after a triumphant attempt to write a book report using the activities he had learned in the reciprocal teaching training, "Mrs. X, you'll be glad to hear this wasn't all for nothing."

All of the teachers indicated that they would add reciprocal teaching to their instructional repertoire, using it with their more capable readers as well as their poor comprehenders. The one eighth-grade teacher planned to divide her entire class of 20 students into four small groups to which she would assign one of the students she had trained in reciprocal teaching. This student would function as group leader in a peer tutoring situation. One of the two remedial reading teachers planned to implement the procedure with her younger students, reading the passages orally, rather than silently.

The students' responses to a post training questionnaire indicated that they also favorably evaluated the procedure — particularly the opportunity to assume the role of teacher. Students, evaluating the procedure, indicated that "finding the good right question" was the most difficult activity and that summarizing was the most helpful activity.

GENERAL DISCUSSION

Summary of Main Findings

These studies can be regarded as successful for eight main reasons: (1) There was clear qualitative evidence of improvement in the students' dialogues. (2) The quantitative improvement on the comprehension tests was large and reliable; all but one student in Study 1 and all the students in Study 2 improved to the level set by good comprehenders. (3) The effect was durable; maintenance probes showed no drop in the level of performance for up to an eight-week period. (4) The effect generalized to the classroom setting, with students reaching or surpassing the average level for their age mates. Given the difficulty reported in obtaining generalization of trained cognitive skills across settings (Brown & Campione, 1981; Meichenbaum, 1977), this is an impressive finding. (5) Training resulted in reliable transfer to laboratory tasks that differed in surface features from the training and assessment tasks — summarizing, predicting questions, and detecting incongruities all improved. Again, this is an impressive finding, given prior difficulty with obtaining

transfer of cognitive skills training (Brown & Campione, 1978, 1981; Brown, Campione, & Day, 1981). (6) Sizable improvements in standardized comprehension scores were recorded for four of the six subjects in Study 1. (7) The intervention was no less successful in natural group settings conducted by regular teachers than it was when conducted by the experimenter. (8) The teachers were uniformly enthusiastic about the procedure once they had mastered it (not before) and planned to incorporate it into their routine teaching repertoires.

Success of Reciprocal Teaching

The success of the reciprocal teaching intervention could be attributed to the particular strategies trained, to the reciprocal teaching procedure, or to a combination of both. In this sense, the studies were multiply confounded. Given the typically limited outcome of cognitive skills training studies (Brown, Bransford, Ferrara, & Campione, 1983), however, we advocate the procedure of *first* obtaining a sizable, durable, and generalized effect and then conducting the necessary research to determine the subcomponents that are primarily responsible for the improvement. And, given the success of the current studies in terms of extent, breadth, and durability of improvement, further research to determine exactly which combination of factors was primarily responsible is now warranted.

The instruction provided during the reciprocal teaching sessions involved extensive modeling and practice in four strategies that were deemed to be ideal comprehension-fostering and comprehension-monitoring activities. Previous attempts to train such activities as summarization (Day, 1980; Linden & Wittrock, 1981) or clarification (Markman & Gorin, 1981) had achieved some success, but the breadth and durability of improvement did not compare with that of the reciprocal teaching interventions reported here. However, both the extent and form of the instruction differed in the current studies. Typically, training studies involve a single instructional session or at most a couple of repetitions of the instruction. In contrast, the studies described here involved 20 days of lessons, continuing over a 4-week period. It remains to be seen whether extensive instruction in the four strategies, alone or in combination, would also succeed as dramatically if the instruction were not of the reciprocal teaching variety (that is, demonstrating alone, direct instruction, etc.). We are currently conducting studies to deal with these issues.

The reciprocal teaching method itself could be the prime reason for success. First, it involves extensive modeling of the type of comprehension-fostering and comprehension-monitoring activities that are usually difficult to detect in the expert reader, as they are executed covertly. The reciprocal teaching procedure provides a relatively natural forum for the teacher to engage in these activities overtly, and hence to provide a model of what it is that expert readers do when they try to understand and remember texts.

Second, the reciprocal teaching routines force the students to respond, even if the level of which they are capable is not yet that of an expert. But because the students do respond, the teacher has an opportunity to gauge their competence and provide appropriate feedback. In this way, the procedure provides an opportunity for the students to make overt their level of competence, a level that in many procedures is masked by students' tendency not to respond until they approach full competence.

Thus, the reciprocal teaching procedure involves continuous trial and error on the part of the student, married to continuous adjustment on the part of the teacher to their current competence. In the current studies, through interaction with the supportive teacher and their more knowledgeable peers, the students were led to perform at an increasingly more mature level; sometimes this progress was fast, sometimes slow, but, irrespective of the rate, the teacher provided an opportunity for the students to respond at a slightly challenging level. The teacher did not merely instruct the students and then leave them to work unaided; she entered into an interaction where the students and the teacher were mutually responsible for getting the task done. As the students adopted more of the essential skills initially undertaken by the adult, the adult acted less as a model and more like a sympathetic coach. In order to perform this essential role, however, the teacher must somehow be sensitive to each student's needs at any stage of the process. She must engage in *on-line diagnosis* that will guide her own level of participation, a level of participation that is finely tuned to the student's changing cognitive status. Diagnosis involves more than initial estimates of starting competence; it also involves continuous evaluation and revision in the teacher's theory of the student's competence, a theory that must be responsive to the level of participation of which the student is currently capable.

The reciprocal teaching training was intended to mimic the conditions of natural learning; every attempt was made to base the reciprocal teaching interaction on the forms of guided learning that are said to occur naturally between experts and novices, both in the ideal home and school settings and in the workplace. The teacher models and explains, relinquishing part of the task to the novices only at the level each one is capable of negotiating at any point in time. Increasingly, as a novice becomes more competent, the teacher increases her demands, requiring participation at a slightly more challenging level, that is, one stage further into the zone of proximal development (see Tables, 4, 5, 6, 11, and 12). We believe that by studying naturally occurring interactive learning situations and conducting training studies that mimic their essential features, we will become better able to understand the conditions that foster learning and, therefore, become more successful at designing effective instruction.

Finally, we would like to argue that training studies can be more than just successful exercises in applied cognition with immediate applicability to school settings. They also afford more direct tests of causal relationships that

cannot be tested within typical correlational approaches. (Campione & Armbruster, in press). Much developmental research is correlational in nature, and there are problems with interpreting such data. To give an example from our own work, in many of our studies, we consider the performance of students who do or do not spontaneously adopt an appropriate text processing strategy; that is often the major variable carrying a developmental trend. For example, fifth and seventh graders who make adequate rough drafts when paraphrasing (Brown, Day, & Jones, 1983), or spontaneously underline or take notes of important text elements while studying (Brown & Smiley, 1978), perform as well as the majority of twelfth graders, whereas twelfth graders who fail to employ these activities look more like fifth graders. This pattern suggests that it is the strategy that leads to efficiency, and developmental trends showing improvement with age are created by the increased proportion of strategic subjects. This is a reasonable interpretation, but as the data are primarily correlational, it cannot be made with certainty. For example, it could be that the young spontaneous strategy users are the more efficient children in general and would perform better than their peers on any task, and on the particular task in question, without the use of strategies. Even partialing out ability factors such as IQ or reading scores does not totally bypass this problem.

One answer to such a dilemma is to resort to converging operations, thus providing a constellation of evidence to buttress a theoretical point. The training study has its place in this constellation. We have advocated a three-pronged approach to providing convergent evidence regarding the importance of any process as an underlying cognitive mechanism (Brown, Palincsar, & Armbruster, in press; Brown & Campione, in press). First, a theory about the underlying processes involved in, for example, reading comprehension, is formulated. Next is the correlational step. Students who read well are also found to perform well on the identified underlying processes, while poor readers experience particular difficulty on just these activities (Brown, Armbruster, & Baker, in press). Finally, students who are not using the specified activities are given training designed to induce the use of the processes theoretically specified as key activities underlying efficiency; others are not. If the theory is correct (and training adequate), trained students' performance should become more like that of spontaneous users, in terms of both levels and patterns (Brown, 1975).

The particular strategies trained in these studies were selected only after a great deal of theoretical discussion about them had occurred (Brown, 1980; Brown & Day, 1983; Collins & Smith, 1982; Flavell, 1981; Kintsch & vanDijk, 1978; Markman, 1981; Stein & Trabasso, 1982) together with numerous studies investigating the correlates of good versus poor reader status (Brown, Armbruster, & Baker, in press). The present work, therefore, represents an instance of the third prong in the overall research strategy. Positive

data from the training studies provide an important source of converging evidence for the theoretical position that contributed to their design. Thus, while there are considerable problems with interpreting the outcomes of training studies (Brown, 1974; Brown & Campione, 1978; Campione & Armbruster, in press), they do provide an important manipulative tool to aid theory development. From the point of view of both theory development and successful cognitive instruction, training studies such as those reported here are valuable tools for enhancing our understanding of the underlying cognitive mechanisms involved in reading and studying.

ACKNOWLEDGMENTS

REFERENCES

Aaron, I. E., Artley, A. S., Goodman, K. S., Jenkins, W. A., Manning, J. C., Monroe, M., Pyle, W. J., Robinson, H. M., Schiller, A., Smith, M. B., Sullivan, L. M., Weintraub, S., & Wepman, J. M. (1976). *Reading unlimited.* Glenview, IL: Scott Foresman.

Allington, R. (1980). Teacher interruption behavior during primary-grade oral reading. *Journal of Educational Psychology, 72*(3), 371–377.

Anderson, R. C. (1978). Schema-directed processes in language comprehension. In A. Lesgold, J. Pellegrino, S. Fokkema & R. Glaser (Eds.), *Cognitive psychology and instruction* (pp. 67–82). New York: Plenum.

Anderson, T. H., & Armbruster, B. B. (in press). Studying. In P. D. Pearson (Ed.), *Handbook of reading research.* New York: Longman.

Andre, M. D. A., & Anderson, T. H. (1978–79). The development and evaluation of a self-questioning study technique. *Reading Research Quarterly, 14,* 605–623.

Au, K. H. (1979). Using the experience-text-relationship method with minority children. *The Reading Teacher, 32,* 677–679.

Backler, A., & Lazarus, S. (1980). *World geography.* Palo Alto, CA: Science Research Associates.

Baker, L. (in press). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest, G. E. Mackinnon & T. G. Waller (Eds.), *Metacognition, cognition and human performance.* New York: Academic Press.

Baker, L., & Brown, A. L. (in press). Metacognitive skills and reading. In D. Pearson, M. Kamil, R. Barr & P. Mosenthal (Eds.), *Handbook of reading research.* New York: Longman.

Bartlett, E. J. (1979). Curriculum, concepts of literacy, and social class. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 2) (pp. 229–241). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bereiter, C., & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2) (pp. 1–64). Hillsdale, NJ: Lawrence Erlbaum Associates.

Binet, A. (1909). *Les idees modernes sur les enfants.* Paris: Ernst Flamarion.

Borkowski, J. G., & Cavanaugh, J. (1979). Maintenance and generalization of skills and strategies by the retarded. In N. R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research* (pp. 190–220). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L. (1974). The role of strategic behavior in retardate memory. In N. R. Ellis (Ed.), *International review of research in mental retardation* (Vol. 7) (pp. 55–111). New York: Academic Press.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10) (pp. 103–152). New York: Academic Press.

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 1) (pp. 77–165). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 453–481). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L. (in press). Mental orthopedics: A conversation with Alfred Binet. In S. Chipman, J. Segal & R. Glaser (Eds.), *Thinking and learning skills: Current research and open questions* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L., Armbruster, B. B., & Baker, L. (in press). The role of metacognition in reading and studying. In J. Orasanu (Ed.), *Reading comprehension: From research to practice.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology* (4th ed.). *Cognitive development* (Vol. 3) (pp. 420–494). New York: Wiley.

Brown, A. L., & Campione, J. C. (1978). Permissible inferences from cognitive training studies in developmental research. In W. S. Hall & M. Cole (Eds.), *Quarterly Newsletter of the Institute for Comparable Human Behavior, 2*(3), 46–53.

Brown, A. L., & Campione, J. C. (1981). Inducing flexible thinking: A problem of access. In M. Friedman, J. P. Das & N. O'Connor (Eds.), *Intelligence and learning* (pp. 515–529). New York: Plenum.

Brown, A. L., & Campione, J. C. (in press). Three faces of transfer: Implications for early competence, individual differences, and instruction. To appear in M. Lamb, A. Brown & B. Rogoff (Eds.), *Advances in developmental psychology* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Brown, A. L., Day, J. D., & Jones, R. S. (1983). The development of plans for summarizing texts. *Child Development, 54,* 968–979.

Brown, A. L., & Palincsar, A. S. (1982). Inducing strategic learning from texts by means of informed, self-control training. *Topics in Learning and Learning Disabilities, 2*(1), 1–17.

Brown, A. L., Palincsar, A. S. & Armbruster, B. B. (in press). Instructing comprehension-fostering activities in interactive learning situations. In H. Mandl, N. Stein & T. Trabasso (Eds.), *Learning from texts.* Hillsdale, NJ: Lawrence Erlbam Associates.

Brown, A. L., Palincsar, A. S., & Purcell, L. (in press). Poor readers: Teach, don't label. In U. Neisser (Ed.), *The academic performance of minority children.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development, 48,* 1–8.

Brown, A. L., & Smiley, S. S. (1978). The development of strategies for studying texts. *Child Development, 49,* 1076–1088.

Brown, A. L., Smiley, S. S., & Lawton, S.C. (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development, 49,* 829–835.

Campione, J. C., & Armbruster, B. B. (in press). An analysis of the outcomes and implications of intervention research. In H. Mandl, N. Stein & T. Trabasso (Eds.), *Learning from texts.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Chipman, S., Segal, J., & Glaser, R. (Eds.) (in press). *Thinking and learning skills: Research and open questions* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clymer, T. (1976). *Reading 720.* Lexington, MA: Ginn & Company.

Collins, J. (1980). Differential treatment in reading groups. In J. Cook-Gumperz (Ed.), *Educational discourse* (pp. 98–115). London: Heinneman.

Collins, A., & Smith, E. E. (1982). Teaching the process of reading comprehension. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 173–185). Norwood, NJ: Ablex.

Collins, A., & Stevens, A. (1982). Goals and strategies of inquiring teachers. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2) (pp. 65–119). Hillsdale, NJ: Lawrence Erlbaum Associates.

Day, J. D. (1980). *Training summarization skills: A comparison of teaching methods.* Unpublished doctoral dissertation, University of Illinois.

Durr, W. K., Pescosolido, J., & Toetter, W. M. (Eds.). (1974). *Serendipity.* New York: Houghton-Mifflin.

Engelman, S., Becker, W. C., Hanner, S., & Johnson, G. (1978). *Correctve reading decoding.* Palo Alto, CA: Science Research Associates.

Flavell, J. H. (1981). Cognitive monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 35–60). New York: Academic Press.

Franks, J. J., Vye, N. J., Auble, P. M., Mezynski, A. J., Perfetto, G. A., Bransford, J. C., Stein, B. S., & Littlefield, J. (1982). Learning from explicit vs. implicit text. *Journal of Experimental Psychology: General, 111*(4), 414–422.

Fry, E. (1977). Fry's readability graph: Clarification, validity, and extension to level 17. *Journal of Reading, 21*(3), 242–252.

Harris, P. L., Kruithof, A., Terwogt, M. M, & Visser, P. (1981). Children's detection and awareness of textual anomaly. *Journal of Experimental Child Psychology, 31,* 212–230.

Heath, S. B. (1981). Questioning at home and at school: A comparative study. In G. Spindler (Ed.), *Doing ethnography: Educational anthropology in action* (pp. 102–131). New York: Holt, Rinehart, & Winston.

Heimler, C. H., & Lockard, J. D. (1977). *Focus on life science.* Columbus, OH: Charles E. Merrill Publishing Co.

Johnson, N. S. (1983). What do you do if you can't tell the whole story? The development of summarization skills. In K. E. Nelson (Ed.), *Children's language* (Vol. 4) (pp. 315–383). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kintsch, W., & vanDijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Larkin, J. H., Heller, J. I., & Greeno, J. G. (1980). Instructional implications of research on problem solving. In W. J. McKeachie (Ed.), *Cognition, college teaching, and student learning* (pp. 51–66). San Francisco: Jossey-Bass.

Linden, J., & Wittrock, M. C. (1981). The teaching of reading comprehension according to the model of generative learning. *Reading Research Quarterly, 16*(1), 44–57.

Locke, E. Q. (1975). *A guide to effective study.* New York: Springer.

Lovitt, T. C., & Hansen, C. L. (1976a). Round one — placing the child in the right reader. *Journal of Learning Disabilities, 6,* 347–353.

Lovitt, T. C., & Hansen, C. L. (1976b). The contingent use of skipping and drilling to improve oral reading and comprehension. *Journal of Learning Disabilities, 9*(8), 486.

Mandler, J. M. (1983). Representation. In J. H. Flavell & E. M. Markman (Eds.), *Carmichael's manual of child psychology* (Vol. 3) (pp. 420–494). New York: Wiley.

Manzo, A. V. (1968). *Improving reading comprehension through reciprocal questioning.* Unpublished doctoral dissertation, Syracuse University.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50,* 643–655.

Markman, E. M. (1981). Comprehension monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 61–84). New York: Academic Press.

Markman, E. M., & Gorin, L. (1981). Children's ability to adjust their standards for evaluating comprehension. *Journal of Educational Psychology, 83*(3), 320–325.

Matteoni, L., Lane, W. H., Ducher, F., & Burns, V. G. (1980). *Keys to reading.* Oklahoma City, OK: The Economy Company.

McDermott, R. P. (1978). Some reasons for focusing on classrooms in reading research. In P. D. Pearson & J. Hansen (Eds.), *Reading: Disciplined inquiry in process and practice* (pp. 213–217). Clemson, SC: National Reading Conference.

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

Meichenbaum, D., & Asarnow, J. (1978). Cognitive behavioral modification and metacognitive development: Implications for the classroom. In P. Kendall & S. Hollon (Eds.), *Cognitive behavioral interventions: Theory, research, and procedure* (pp. 11–35). New York: Academic Press.

Nieman, E. W., & Safier, F. (1979). *Adventures for readers.* New York: Harcourt Brace Jovanovich.

Ninio, A., & Bruner, J. S. (1978). The achievement and antecedents of labeling. *Journal of Child Language, 5,* 1–15.

Palincsar, A. S. (1981). *Strategy training versus corrective feedback in the instruction of reading comprehension.* Unpublished paper, University of Illinois.

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.

Pearson, P. D., & Johnson, D. D. (1978). *Teaching reading comprehension.* New York: Holt, Rinehart & Winston.

Raphael, T. E. (1980). *The effects of metacognitive strategy awareness training on students' question answering behavior.* Unpublished doctoral dissertation, University of Illinois.

Raphael, T. E., Wonnocat, C. A., & Pearson, P. D. (1983). *Increasing students' sensitivity to sources of information: An instructional study in question-answer relationships* (Tech. Rep. No. 284). Urbana: University of Illinois, Center for the Study of Reading.

Rogoff, B., & Gardner, W. (in press). Guidance in cognitive development: An examination of mother-child instruction. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context.* Cambridge, MA: Harvard University Press.

Stein, N. L., & Trabasso, T. (1982). What's in a story: Critical issues in comprehension and instruction. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2). (pp. 271–282). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes,* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman, Eds. and trans.). Cambridge, MA: Harvard University Press.

Wertsch, J. V., & Stone, C. A. (1979, February). *A social interactional analysis of learning disabilities remediation.* Paper presented at the International Conference of the Association for Children with Learning Disabilities. San Francisco, CA.

Wong, B., & Jones, W. (1981). *Increasing metacomprehension in L. D. normally-achieving students through self-questioning training.* Unpublished manuscript, Simon Fraser University.

Wood, D., & Middleton, D. (1975). A study of assisted problem-solving. *British Journal of Psychology, 66,* 181–191.