Instructing Comprehension-Fostering Activities in Interactive Learning Situations

Ann L. Brown, University of California, Berkeley

Annemarie Sullivan Palincsar, University of Michigan

Bonnie B. Armbruster, University of Illinois at Urbana-Champaign

Introduction

“Pupils don’t learn to comprehend by osmosis” (Cushenbery, 1969). As with any definite statement concerning the acquisition of reading skills, this could be a controversial position, but a main theme of this piece is that, at least for a sizable number of children, the statement is true. It is also argued that children who need extensive instruction in comprehending written materials most are least likely to receive it. The latter part of the piece describes cognitive-skills training studies that have provided extensive practice in comprehension-fostering activities and have resulted in substantial improvements in students’ ability to learn from texts.

Resnick (1979) has argued that there are two main biases in reading instruction, namely direct instruction of decoding and informal teaching of comprehension. Those who advocate a heavy emphasis on decoding mechanisms in early reading also tend toward the direct-instruction approach, whereas those who emphasize early attention to language processing, language arts, or comprehension tend also to espouse learner-directed, informal instructional approaches. As Resnick also argued, there is no reason in principle why one cannot have direct instruction in comprehension or (a little harder to envisage) informal instruction in decoding. In this piece, concern is with one of the underpopulated cells, relatively direct or explicit instruction in comprehension. Of particular concern is the explicit instruction of comprehension-fostering skills with children at risk for academic failure precisely because they experience unusual difficulties in this arena.

Prereading Experiences

Preschool Reading Dyads

Learning to read does not begin when the child enters school; the child brings a history of preschool learning experiences that, to a greater or lesser extent, have prepared the way for a smooth transition. Some of these experiences could clearly be classified as prereading activities; others are more general learning practices
with some relevance to reading. Many of these early experiences have taken place in social settings that share pertinent features with common school learning activities. Some children have considerable preschool experience in interactions that are very similar to school reading groups; others have not.

Certain parent-child interactions are ideal practicing grounds for subsequent teacher-child activities that will be of central importance in the early grades. Social settings such as these, where the child interacts with experts in a problem-solving domain, are settings where a great deal of learning occurs in and out of school. Indeed, some would argue that the majority of learning is shaped by social processes (Laboratory of Comparative Human Cognition, 1986; Vygotsky, 1934/1978). From this perspective it is claimed that children first experience a particular set of problem-solving activities in the presence of others and only gradually come to perform these functions for themselves. First, the adult (parent, teacher, etc.) guides the child’s activity, doing much of the cognitive work herself, but gradually the adult and child come to share the cognitive functions with the child taking initiative and the adult correcting and guiding where the child falters. Finally, the adult allows the child to take over the major thinking role and adopts the stance of a supportive and sympathetic audience.

This developmental progress from social to individual cognitive processing (other regulation to self-regulation) is nicely illustrated in parent-child learning dyads such as those reported by Wertsch (1978). Of particular interest to the argument here are the interactions of mothers and children as they engage in picture book “reading.”

At least in middle class homes, a stable locus of parent-child interactions is the picture book task. Ninio and Bruner (1978) observed one mother-infant dyad longitudinally, starting when the child was only 8 months old and terminating (unfortunately) when he was 18 months old. From the very beginning, their interaction can best be described as a dialogue with the timing of mother’s and child’s behavior following an almost complete alternation pattern strikingly similar to the turn-taking conventions observed in dialogue. The mother initially is very much in command and seduces the child into the ritual dialogue for picture book reading by accepting any response from the baby as appropriate for his turn in the conversation. Indeed, Ninio and Bruner point out that the mother accepts an astonishing variety of responses as acceptable turn-taking behavior interpreting anything as having a “specific, intelligible content.” The “imputation of intent and content” to the child’s activities constitutes “an important mechanism by which the child is advanced to more adult-like communicative behavior” (Ninio & Bruner, 1978, p. 8).

A dramatic shift in responsibility comes when the child begins to label pictures for himself. Now the mother acts as if she believes the child has uttered words rather than babble. As the mother’s theory of the child changes, so does her part in the dialogue. At first she appears to be content with any vocalization, but as soon as actual words can be produced the mother steps up her demands and asks for a label with the query “What’s that?” The mother seems to increase her
level of expectation, first “coaxing the child to substitute a vocalization for a non-vocal sign and later a well-formed word for a babbled vocalization.” Initially, the mother does all the labeling because she assumes that the child cannot. According to Ninio and Bruner (1978),

Later, the mother starts a cycle with a label ONLY if she thinks that the child will not label the picture himself, either because he does not yet know the correct word or he is not attentive enough to make the effort at labeling. If circumstances seem more favorable for labeling to occur, she will usually start the cycle with a “What's that?” question. (p. 14)

Responsibility for labeling is transferred from the mother to the child in response to his increasing store of knowledge, finely monitored by the mother. During the course of the study, the mother constantly updated her inventory of the words the child had previously understood and repeatedly attempted to make contact with his growing knowledge base. For example:

1. You haven't seen one of those; that’s a goose.
2. You don't really know what those are, do you? They are mittens; wrong time of year for those.
3. It’s a dog; I know you know that one.
4. We'll find you something you know very well.
5. Come on, you’ve learned “bricks.”

DeLoache (1983) has repeated many of these observations in a cross-sectional study of mothers reading to their children. The children ranged from 17–38 months. The mothers of the youngest children point to the objects and label them, sometimes providing some additional information. In the middle age group, the children are much more active. Their mothers ask them to point to and label objects to provide other information about the picture. These children often spontaneously provide labels (“There's a horsie”) or ask the mothers for labels (“What's this?”). In the oldest group studies, more complex stories were introduced, and the mothers do much more than simply label objects. They talk about the relation among the objects in the picture and relate them to the child’s experience—e.g., “That's right, that's a bee hive. Do you know what bees make? They make honey. They get nectar from flowers and use it to make honey, and then they put the honey in the bee hive.” When the child can recognize the pictures and knows something about them, the mother uses the material to provide the child with a great deal of background information only loosely related to the actual pictures. It is not simply that the amount of help changes as the child becomes more competent, but the quality of help is finely geared to the child's current level.

In both the Ninio and Bruner and DeLoache dyads, the mother is seen functioning repeatedly in the child's “region of sensitivity to instruction” (Wood & Middleton, 1975) or “zone of proximal development” (Vygotsky, 1934/1978). As the child advances so does the level of collaboration demanded by the mother.
The mother systematically shapes their joint experiences in such a way that the child will be drawn into taking more and more responsibility for the dyad’s work. In so doing, the mother not only provides an optimal learning environment, she also models appropriate comprehension-fostering activities such as elaboration, activating appropriate background knowledge, and questioning strategies. These crucial activities are thereby made overt and explicit.

**Inadequate Early Mediated Learning**

It has been argued that parent-child interactions such as the social reading experiences just described are important preparations for early school success. It has also been argued that a severe lack of interactive experiences is a primary source of academic retardation. A leading advocate of this position is Feuerstein (1979, 1980), who argues that cognitive growth is very heavily dependent on the quality of mediated learning that the child experiences. According to Feuerstein (1979), “Mediated learning is the training given to the human organism by an experienced adult who frames, selects, focuses, and feeds back an environmental experience in such a way to create appropriate learning sets” (p. 6). These mediated-learning experiences are an essential aspect of development beginning when the parent selects significant objects for the infant to focus on and proceeding throughout the development with the adult systematically shaping the child’s learning experiences. Feuerstein believes that this is the principal means by which children develop the cognitive operations necessary for learning independently. By interacting with an adult, who models and guides problem-solving activities and structures learning environments, children gradually come to adopt structuring and regulating activities of their own. It is argued that cognitive skills, including those of comprehension fostering and monitoring, develop normally via a process whereby the adult models and prompts their use and the children gradually adopt such activities as part of their own repertoire.

Feuerstein believes that the principal reason for the poor academic performance of many disadvantaged students is the lack of consistent instruction by parental models in their earlier developmental histories because of parental apathy, ignorance, or overcommitment. Quite simply, parents in disadvantaged homes were often themselves disadvantaged children and cannot be expected to teach what they perhaps do not know; large family size and the need for a working mother do not leave a great deal of time for Socratic dialogue games. In addition, interactive styles of continually questioning and extending the limits of knowledge that are typical of middle class social-interaction patterns (Ninio & Bruner, 1978) may even be alien to some cultures (Au, 1979; Bernstein, 1971).

Mediated-learning activities, however, are exactly what occurs in schools, and the middle class child comes well prepared to take part in these rituals. Not only does the disadvantaged child lack sufficient prior exposure, but there is evidence that teachers give less experience in this learning mode to those who, because of their lack of prior experience, need it most.
Instructing Comprehension-Fostering Activities in Interactive Learning Situations

School Reading Experiences

Teachers as Models of Reading Strategies

Ideally when the child reaches school, teachers take over some of the mediating functions, acting as models and promoters of comprehension-fostering activities. In schools, effective teachers are those who engage in continued prompts to get children to plan and monitor their own reading activities. Effective teachers model many forms of critical thinking for their students (Collins & Stevens, 1982). Thus, Collins and Smith's (1982) recent call for teachers to model comprehension-monitoring activities is timely but not novel. In a recent review of the literature, Bird (1980) traces the history of this idea and points out that it is a recurrent theme in the reading-education literature. For example, Lorge (1957) stressed that the teacher should model active-comprehension processes and direct the child's attention to the thinking and reasoning processes that he or she engages in while reading for meaning. Smith (1961) also called for teachers to ask questions that stimulate the drawing of inferences, the making of predictions and of comparisons, and the use of cause and effect reasoning. This theme was repeated by King (1967) and Gantt (1970), who described programs of teacher-directed questioning by which children are led through the sequence of thinking necessary to understand the passage. Similarly, Schwartz and Scheff (1975) suggested that teachers encourage active reading by demonstrating their own curiosity, posing questions, reasoning, predicting, and verifying inferences and conclusions.

Good teachers do this, of course. Even teachers who refrain from such activities in reading group do demonstrate them in, for example, reading a story to the whole class (Griffin, 1977). In their discussion of why teachers are easier to understand than texts, Schallert and Kleiman (1979) identified four main activities that teachers use to help children understand. They tailor the message to the child's level, activate relevant background knowledge, focus student attention on main points, and force comprehension monitoring by probing and question-asking aimed at testing the degree of understanding.

The main theme of all this work is that the ideal teacher functions as a model of comprehension-fostering and comprehension-monitoring activities largely by activating relevant knowledge and questioning basic assumptions. These are the essential features of the teaching style referred to variously as Socratic, case, or inquiry methods. Collins and Stevens (1982) have examined a variety of teachers and developed a taxonomy of tactics that are commonly used by outstanding teachers, notably the “entrapment ploys” of counterexamples and invidious generalizations, the “extension ploys” that force students to apply their newfound knowledge broadly, and the “debugging ploys” that force students to correct their misconceptions (Collins & Stevens, 1982).

Collins and Stevens point out that a main goal of such dialogues is not to convey the content of a particular domain. If this were the aim, the method would be inefficient due to the low rate of information transfer; more points can be conveyed in a lecture than can be discussed in a Socratic dialogue. If the method is successful, it is because it teaches students to think scientifically, to make
predictions, to question and evaluate. An effective aid to knowledge-building and revision is the ploy of forcing learners to make their theories explicit and to defend them to others.

In order for these activities of questioning, predicting, hypothesis generation, testing, and revision to be of service to the child, it is necessary that they are transferred from the teacher to the child in such a way that they form part of the learner’s battery of comprehension-fostering skills. A common problem with all these approaches, as pointed out by Jenkinson (1969) and Gall (1970), is that they presuppose that children witnessing these activities will come to employ them on their own. This is the problem of internalization, how the child comes to personally use activities that were originally social (Vygotsky, 1934/1978). We return to this point later.

**Reading Groups and Reading Status**

Not all children are exposed to gifted teachers, and even the same teachers may offer different learning environments to those viewed as good or poor readers (Brown, Palincsar, & Purcell, 1986). The selection of a curriculum is itself the selection of a particular reading environment. Bartlett (1979) claimed that disadvantaged children are most likely to be exposed to early reading programs with a heavy emphasis on decoding such as Distar, whereas middle class children receive earlier exposure to programs that emphasize comprehension such as Open Court. Bartlett goes on to compare the types of questions featured prominently in Distar exercises to those recommended by Open Court. In general, the Distar questions tend to focus on locating and remembering specific information, whereas the Open Court questions promote reflection upon and the questioning of the meaning of the text. Hence, a different type of instruction is aimed at children who enter school differentially prepared for the experience.

Even if the curriculum and classroom placement do not differ, there remains evidence that the reading environment is not equal for all children. For example, detailed observations of reading groups (Allington, 1980; Au, 1980; Cazden, 1979; Collins, 1980; McDermott, 1978) have shown that good and poor readers are not treated equally. Good readers are questioned about the meaning behind what they are reading, and they are asked to evaluate and criticize material frequently. A considerable amount of time in the good reading group is “on task”; i.e., reading-related activities occur, and a sizable amount of the group activities are of an optimal “comprehension-fostering” type. In the good reading group, the teacher adopts the procedure of asking every child to read in turn; but in the poorer reading group, turn-taking is at the teacher’s request, and the really poor readers are not called upon to perform to save everyone embarrassment (McDermott, 1978). Precious little time in the poor reading group is spent doing comprehension-fostering activities; the lion’s share of activities involve the establishment of such rituals as turn-taking and hand raising. When and if they are required to read, poor readers receive primarily drill in pronunciation and decoding. Rarely are they given practice in qualifying and evaluating their comprehension (Allington,
In instructing comprehension-fostering activities in interactive learning situations (McDermott, 1978), children who come to school inadequately prepared for reading, for whatever reason, tend to end up in the bottom reading groups and are, therefore, exposed to different reading experiences. The emphasis is clearly on decoding and not on comprehension. If as a result of their initial failure and subsequent treatment these children are singled out for special education, they run the risk of an intensive version of this same treatment; for special education in reading problems has an even heavier emphasis on decoding skills at the expense of reading-comprehension instruction.

A strong emphasis on direct instruction in basic skills permeates resource rooms and special education classrooms, perhaps an understandable reaction to the “lack” of these skills demonstrated by the students. Special education classes are more likely to provide step-by-step instruction for students in basic skills (decoding, etc.) and rarely allow the students to figure out meanings or question their assumptions. Heavily programmed and guided learning of this type may be a practical and efficient means of getting less-successful students to perform better on a particular task, i.e., word recognition. But it is the teacher, not the child, who is making all the learning decisions. Such experience is less likely to be the appropriate procedure for promoting insightful learning. Students may learn something about a particular task, but they are less likely to learn how to learn from reading (Brown, 1982).

Of course we are not arguing against the practice of direct instruction in decoding per se. As Resnick (1979) has pointed out, there is a great deal of evidence to support the success of reading programs that favor early direct instruction of decoding. We do argue, however, that instruction in comprehension can and should be offered in addition (Tharp, 1982), because the current state of affairs is that poor readers, particularly those labeled as learning disabled or mildly retarded, are unlikely in the present system to develop adequate reading-comprehension skills. Decoding is mastered eventually but reading-comprehension scores remain low and possibly permanently and severely depressed. Many factors may be responsible for this typical pattern, but one that is rarely addressed is the simple explanation of practice. Practice makes possible; if so, perhaps we should not be surprised to find a cumulative deficit in comprehension skills in those who are systematically denied extensive experience in comprehension-fostering activities.

In summary, following repeated experience with experts (parents, caretakers, teachers, etc.) who situate, elaborate, evaluate, and extend the limits of their experience, many students develop a battery of school-relevant autocritical skills (Binet, 1909; Brown, 1985) that include comprehension-fostering activities ideally tailored for reading. These skills are essential acquisitions, if students are to learn how to learn independently. If for some reason the child is deprived of a constant history of such interaction in and out of school, the development of an
adequate battery of self-regulatory skills for performing independently on academic tasks may be impeded.

Given this argument, an appropriate training experience would be to attempt to mimic naturally occurring interactive-learning settings as a context for instruction. In a subsequent section, a series of experiments is reviewed that may have promise for improving comprehension skills precisely because they attempt to help children adopt for themselves questioning and monitoring activities that they experience initially in interactive settings.

**Comprehension-Fostering Activities**

Before proceeding to a discussion of instruction, an attempt is made to be somewhat more explicit about the nature of the comprehension processes involved in effective reading. We concentrate on those that promote comprehension and lead to effective comprehension monitoring, i.e., activities engaged in by readers to ensure that comprehension is proceeding smoothly. Although far from a detailed task analysis of reading comprehension, there are several overlapping skills that have been mentioned repeatedly as prime comprehension-fostering activities in a variety of recent theoretical treatments (Baker & Brown, 1984a, 1984b; Brown, 1980; Collins & Smith, 1982; Dansereau, 1980; Markman, 1981). These activities include

1. clarifying the purposes of reading, i.e., understanding the task demands, 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.

All of these activities appear as academic tasks in their own right; for example, it is a common practice to call on children to concentrate on the main idea, to think critically about the content of what they are reading, or to summarize or answer questions on a passage. But, in addition, these activities, if engaged in while reading, serve to enhance comprehension and afford 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-fostering and 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 is being read, 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 (André & Anderson, 1978/1979). Thus, closing one’s eyes (metaphorically) and attempting to state the gist of what one has read and asking questions of an interpretive and predictive nature (Collins & Smith, 1982) 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; the need for explicit instruction in comprehension-enhancing activities is particularly acute for the slow-learning student (Brown & Palincsar, 1982).

Instructing Reading Comprehension

Teaching Settings and Reading Strategies

In this section a series of successful training studies is described which attempt to combine a knowledge of effective teaching settings and appropriate learning activities in order to improve comprehension (Palincsar & Brown, 1983). We have argued that many students lack sufficient practice in interactive-learning situations where comprehension-fostering activities are modeled and promoted. If this were true, then an obvious compensatory strategy would be to design instruction where practice in the essential skills is embedded within an interactive-learning situation that mimics the idealized mother-child, teacher-child dialogues previously described.

The particular skills selected for training were summarizing (self-review), questioning, clarifying, and predicting. There is a large literature connected with each activity. A considerable amount is known about the use or nonuse of the activities in isolation, especially in response to direct instruction. But considerably less is known about the spontaneous orchestration of a battery of such activities in the face of different forms of comprehension failure. For example, high school and junior college students have a great deal of trouble writing adequate synopses of texts (Brown & Day, 1983), although well-designed training can improve these skills (Day, 1980). Very little is known, however, about the use of self-directed paraphrasing as a method of periodic review or as a means of monitoring comprehension when the text gets difficult. Also, young and poor readers have difficulty evaluating texts for clarity, internal consistency, or compatibility with known facts (Markman, 1981), and training helps here too (Markman & Gorin, 1981). But again, little is known about where and when and with what actual processes the reader will engage in such monitoring spontaneously. Similarly, students often fail to generate questions and can be trained to perform better on these skills in isolation (André & Anderson, 1978/1979), but little is known about the spontaneous use of questioning as part of a concerted, personally
designed, and coordinated plan of attack in the face of comprehensive difficulties. Therefore, in this series of studies the four activities of self-directed summarizing (review), questioning, clarifying, and predicting are combined in a package of activities with the general aim of enhancing understanding. Each “separate” activity, however, was used in response to a concrete problem of text comprehension. Clarifying occurred only if there were confusions either in the text (unclear referent, etc.) or in the student’s interpretation of the text. Summarizing was modeled as an activity of self-review; it was engaged in in order to state to the group what had just happened in the text and as a 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. Questioning, similarly, was not practiced as a teacher-directed isolated activity, but as a concrete task—what question would a teacher or test ask about that section of the text. Students reacted very positively to this concrete detective work, rather than the more typical isolated skills-training approach as we will see.

We embedded these activities within a training procedure that was very similar to the interactive mother-child, teacher-student dyads described earlier. The procedure was also similar to that of reciprocal questioning. Manzo (1969) introduced a variant of this with his ReQuest procedure. Teachers and small groups of remedial-reading students took turns asking themselves questions about what they were reading. Questions followed every sentence, a procedure that would not encourage synthesis across larger segments of text. And the types of questions modeled and generated were not necessarily optimal. For example, one teacher modeled the question, “What was the third word in the first sentence?” Even so, Manzo reported significant improvement in standardized reading-comprehension scores. Frase and Schwartz (1975) also had college students taking turns generating or answering questions. Regardless of which role the students assumed, they performed better than when engaged in silent reading. Even though training was not extensive and again there was no attempt to ensure adequate quality of questions, the intervention produced a modest but reliable effect. Given these promising precursors, a reciprocal teaching method was adopted where, in addition to question generation, the activities of reciprocal paraphrasing, clarifying, and predicting were added.

**Instructing Comprehension-Fostering by Reciprocal Teaching**

So far three studies have been completed (Palincsar & Brown, 1983). The first two are laboratory studies with an experimenter (Palincsar) interacting with individual children or with pairs of students. The third study was conducted in the classroom by regular classroom teachers. The students in all three studies were seventh graders with average decoding skills but seriously deficient comprehension scores.
Study 1. In Study 1, four students served as subjects in an extensive training experiment (for full details see Palincsar & Brown, 1983). Each subject served as his or her own control. After completing the decoding and comprehension tests that made them eligible for the study, the students received a period of baseline assessment, on each day of which they read a 500-word expository passage and then attempted to answer 10 comprehension questions independently. This baseline procedure of reading and answering questions on a novel assessment passage each day was also followed during maintenance and long-term follow-up periods. During training periods the students also read and answered questions on a novel assessment passage, but the assessment stage was preceded by interactive training sessions on still different passages. All data reported are percent correct from the daily independent-assessment test, not from the interacted-upon texts.

There were 6–8 days of initial baseline, 10 days of reciprocal teaching, followed by 6 days of maintenance and then a further 3 days of reciprocal teaching. Six months later, the students were retested for 8 days—4 days of untreated maintenance followed by 2 days where reciprocal teaching was reintroduced, which was followed in turn by a final 2 days of maintenance.²

During the reciprocal-teaching intervention, the investigator and the student engaged in an interactive-learning game that involved taking turns in leading a dialogue concerning each segment of text. If the passage was new, the investigator called the student's attention to the title, asked for predictions based upon the title, and discussed the relationship of the passage to prior knowledge. For example, if the passage were entitled Ship of the Desert, the investigator and student would speculate about what the passage might concern and would review what they knew about the characteristics of the desert. If the passage were partially completed, the investigator would ask the student to recall and state the topic of the text and several important points already covered in the passage.

The investigator then assigned a segment of the passage to be read (usually a paragraph) and either indicated that it was her turn to be the teacher or assigned the student to teach that segment. The investigator and student then read the assigned segment silently. After reading the text, the teacher for that segment summarized the content, discussed and clarified any difficulties, asked a question that a teacher or test might ask on the segment, 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.

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 that the questions will be on a test are sure ways of testing oneself to see if one has understood.

At first the students had difficulty taking their part in the dialogue, experiencing particular difficulties with summarizing and formulating questions. The adult teacher helped with a variety of prompting techniques such as “What
question did you think a teacher might ask?” “Remember, a summary is a shortened version; it doesn't include detail” and “If you're having a hard time summarizing, why don't you think of a question 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 that continued to need improvement: “A question I would have asked would be...”; “I would summarize by saying...”; “Did you find this statement unclear?”

Initially, then, the experimenter modeled appropriate activities, but the students had great difficulty assuming the role of dialogue leader when their turn came. The experimenter was sometimes forced to resort to constructing paraphrases and questions for the student to mimic. In this initial phase, the experimenter was modeling effective comprehension-monitoring strategies, but the student was a relatively passive observer.

In the intermediate phase, the students became much more capable of playing their role as dialogue leader and by the end of 10 sessions were providing paraphrases and questions of some sophistication. For example, in the initial sessions, 46% of questions produced by the students were judged as nonquestions or as needing clarification. By the end of the sessions only 2% of responses were judged as either needing clarification or nonquestions. Unclear questions drop out and are replaced over time with questions focusing on the main idea of each text segment. Examples of questions judged to be needing clarification, main idea, and detail are shown in Table 1.

A similar improvement was found for summary statements. At the beginning of the session, only 11% of summary statements captured main ideas, whereas at the end 60% of the statements were so classified. Examples of summary statements are shown in Table 2.

With repeated interaction with a 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 was constructed by integrating information presented across several sentences. Given the steady improvement on the privately read texts, it would appear that students internalize these activities as part of their own repertoire of
comprehension-fostering skills. In support of this statement are the data from peer tutoring sessions taken at the termination of the study. Trained tutees faced with naive peers did attempt to model main idea paraphrase and questions (Palincsar & Brown, work in progress).

In addition to the qualitative changes in the students’ dialogues, there was a gratifying improvement in the level of performance on the daily assessment

Table 1. Examples of Student-Generated Questions During Reciprocal Teaching

<table>
<thead>
<tr>
<th>Main-Idea Questions</th>
<th>Questions Pertaining to Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why don't people live in the desert?</td>
<td>How far south do the maple trees grow?</td>
</tr>
<tr>
<td>Why are the grasslands of Australia ideal for grazing?</td>
<td>What color is the guards’ uniforms?</td>
</tr>
<tr>
<td>What does the light on the fish do?</td>
<td>How many years did it take to build the Great Wall?</td>
</tr>
<tr>
<td>What did these people [the Chinese] invent?</td>
<td>What are chopsticks made out of?</td>
</tr>
<tr>
<td>Plans are being made to use nuclear power for what?</td>
<td>Tell me where the cats hide?</td>
</tr>
<tr>
<td>What are three main problems with all submarines?</td>
<td>What was the balloon material made of?</td>
</tr>
<tr>
<td>Is there just one kind of explosive?</td>
<td>What (on the fish) overlaps like shingles on a roof?</td>
</tr>
<tr>
<td>What are one of the three things people used explosives for?</td>
<td>How far can flying fish leap?</td>
</tr>
<tr>
<td>What are the Philippine officials going to do for the people?</td>
<td>What is the temperature along the southern shores of Australia?</td>
</tr>
</tbody>
</table>

Questions Requiring Clarification (and Suggested Appropriate Questions Regarding the Same Material 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 does it keep the ground? (What effect does snow have on the ground?) |
| What were some of the people? (What kinds of people can serve on a jury?) | 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 kinds of thing? Wait a minute. OK. What was the Manaus tree built for? (Why was the city of Manaus built?) | 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?) |
| 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. OK? (Name some explosives.) |
| 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. OK? (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?) |
Brown, Palincsar, and Armbruster

question-answering score. The students averaged 15% correct during baseline. After the introduction of the reciprocal teaching, the students reached accuracy levels of 80–90% correct. This level was durable across both the maintenance and brief reintroduction of the intervention. After the 6-month delay, the students

<table>
<thead>
<tr>
<th>Main-Idea Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>It says if a man does his job real good, then he will do better in his next life.</td>
</tr>
<tr>
<td>I learned that they have different kinds of Gods, not just Brahman, every family has their own.</td>
</tr>
<tr>
<td>It tells us about the two kinds of camels, what they are like and where they live.</td>
</tr>
<tr>
<td>My summary is that the part of the earth that we live on and see and know is the top layer, the crust.</td>
</tr>
<tr>
<td>This paragraph talks about what happens when people perspire or sweat. They lose a large amount of salt and they get weakness.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is a pair of fins which look like legs.</td>
</tr>
<tr>
<td>The sea horse always swims head up.</td>
</tr>
<tr>
<td>There were large lizards and four eyed fish and 30 foot dandelion.</td>
</tr>
<tr>
<td>What I learned is that a submarine went around the world in 84 days.</td>
</tr>
<tr>
<td>I learned that Cousteau’s first artificial island was in the North Sea.</td>
</tr>
<tr>
<td>Professor Charles went 27 miles and rose 2000 feet in his balloon.</td>
</tr>
<tr>
<td>They [the aborigines] don’t wear much clothes on.</td>
</tr>
<tr>
<td>They [Egyptians] made bread a long time ago.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incomplete Statements and Corresponding Text Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>They talk about it was the richest island; but it didn’t have something, OK, 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.)</td>
</tr>
<tr>
<td>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.)</td>
</tr>
<tr>
<td>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.)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Examples of Student-Generated Critical/Evaluative Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Boy, the paragraph sure is a mess. It is all over the place.”</td>
</tr>
<tr>
<td>“I don’t see how they can say ‘heat lightning occurs on hot summer days.’ How could you see it?”</td>
</tr>
<tr>
<td>“It says here ‘cloud to cloud’ then ‘cloud to earth.’ Wouldn’t that be the same thing?”</td>
</tr>
<tr>
<td>“The word meter throws me off in this sentence.”</td>
</tr>
<tr>
<td>“What’s the difference between soap and detergent anyway?”</td>
</tr>
<tr>
<td>“At first I didn’t get this because I thought the word pumping was bumping.”</td>
</tr>
<tr>
<td>“I don’t know what omitting is.”</td>
</tr>
<tr>
<td>“I have one, what do they mean by ‘far away dreams?’”</td>
</tr>
</tbody>
</table>
averaged 60% correct without help, significant savings over their original level of 15%. After only 1 day of renewed reciprocal teaching, the performance of 2 students returned to 80% and for the remaining students it reached 90% correct; again, the levels were maintained when the intervention was removed. Remember that these scores were obtained on the privately read assessment passages, i.e., 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.

**Generalization to classroom settings.** Throughout the study, a series of 5 probes was made in the social studies classroom setting to see if the students would show any improvement on the identical task of answering 10 comprehension questions on a test. The students were not told that these tests, administered by the classroom teacher, had anything to do with the intervention. All seventh graders took the social studies test as part of their regular classroom activity. The experimental students began the study below the 15th percentile on this task compared with the remaining seventh graders in their school. Performance fluctuated widely, which was not surprising as little was done to promote generalization to the classroom; e.g., the classroom teaching did not encourage the use of strategies and the students received no feedback regarding classroom performance. However, the following mean gains in percentile ranks were obtained between the baseline and final probes: Student 1 = 20, Student 2 = 46, Student 3 = 4, and Student 4 = 34.

In summary, students in Study 1 showed a dramatic improvement in their ability to answer comprehension questions on independently read texts. This improvement was durable in the resource room setting and showed some tendency to generalize to the classroom setting. In addition, qualitative improvement in the students’ dialogues reflected their increasing tendency to concentrate on questions and summaries of the main idea. The reciprocal-teaching procedure was a powerful intervention for improving comprehension.

**Study 2.** Encouraged by the success of the initial study, it was decided to replicate the main features of the successful reciprocal-teaching procedure with 6 additional students, in 3 groups of 2. In addition to group size, the second study also differed from the first in that (1) a criterion level of 75% correct on 4 out of 5 consecutive days was established; (2) students received explicit (graphed) knowledge of results; and (3) tests of transfer were included.

The tests of transfer were selected because it was 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. 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. In addition, two other tests were used as measures of general comprehension monitoring, error detection (Harris, Kruithof, Terwogt,

There were 4 phases to the study. As in Study 1, each student was given a daily assessment passage on which he or she answered 10 comprehension questions, and this was all that occurred on baseline and maintenance days. On intervention days, the assessment passage was preceded by the reciprocal-teaching intervention, identical to that described in Study 1. The phases of Study 2 were as follows:

1. Variable baseline consisting of 4 days for Group 1, 6 days for Group 2, and 8 days for Group 3
2. Reciprocal-teaching intervention consisting of approximately 20 days
3. Maintenance consisting of 5 days of testing at the termination of training
4. Long-term follow-up that took place 8 weeks later (3 days)

All students were appraised of their progress on a daily basis. They were shown graphs depicting the percentage correct for the previous day’s assessment.

The data from the daily assessment passages are shown in Figure 1. The 6 students of Study 2 had baseline accuracy not exceeding 40% correct. They proceeded to make stepwise progression toward means in excess of 75%. Four of the six students reached a stable level of 80% for 5 successive days, taking 12, 11, 11, and 15 days respectively to do it (Students 1, 3, 4, and 6). Student 5 reached criterion of 75% correct in 12 days. Student 2 was the only “failure”; she progressed from a baseline of 12% correct and reached a steady level of 50% correct in 12 days, a significant improvement, but she never approached the criterion level of the remaining 5 students. All students maintained their improved level of performance on both short- and long-term maintenance tests.

A similar improvement in the quality of the dialogues over time was found in Study 1 and Study 2 (see Palincsar & Brown, 1983, for details). At the outset, students required more assistance with the dialogue, asked more unclear and detailed questions, and made more incomplete/incorrect or detailed summaries than they did on the last intervention day. Both main-idea questions and paraphrases increased significantly over time.

Students improved at differential rates. For example, Student 6, a minority student whose Slossen test indicated an IQ of 70, made steady but slow progress as indicated by the dialogue shown in Table 3. The data are taken from Days 1 to 15, the day on which he reached criterion. From a very slow start, this student did achieve an acceptable level of performance both on the dialogues and on his daily assessment passages.

Generalization probes taken in the classroom setting resulted in variable performance but did show clear evidence of improvement. Probes were taken in two settings, social studies and science. At baseline on the social studies probe the range of percentile rankings was .9–43, with four students at or below the 5th percentile. The percentile rankings were typically higher in science with a range
Figure 1. An Example of the Daily Data From Individual Subjects During Baseline, Intervention, Maintenance, and Long-Term Follow-Up

These data are taken from the six experimental subjects of Palincsar and Brown (1983), Study 2.
Table 3. The Acquisition of Question-Asking by One Seventh-Grade Student

<table>
<thead>
<tr>
<th>Day 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>S:  What is found in the southeastern snakes, also the copperhead, rattlesnakes, vipers—they have. I’m not doing this right.</td>
</tr>
<tr>
<td>T:  All right. Do you want to know about the pit vipers?</td>
</tr>
<tr>
<td>S:  Yeah.</td>
</tr>
<tr>
<td>T:  What would be a good question about the pit vipers that starts with the word why?</td>
</tr>
<tr>
<td>S:  (no response)</td>
</tr>
<tr>
<td>T:  How about, “Why are the snakes called pit vipers?”</td>
</tr>
<tr>
<td>S:  Why do they want to know that they are called pit vipers?</td>
</tr>
<tr>
<td>T:  Try it again.</td>
</tr>
<tr>
<td>S:  Why do they, pit vipers in a pit?</td>
</tr>
<tr>
<td>T:  How about, “Why do they call the snakes pit vipers?”</td>
</tr>
<tr>
<td>S:  Why do they call the snakes pit vipers?</td>
</tr>
<tr>
<td>T:  There you go! Good for you.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>S:  (no question)</td>
</tr>
<tr>
<td>T:  What’s this paragraph about?</td>
</tr>
<tr>
<td>S:  Spinner’s mate. How do spinner’s mate....</td>
</tr>
<tr>
<td>T:  That’s good. Keep going.</td>
</tr>
<tr>
<td>S:  How do spinner’s mate is small than.... How am I going to say that?</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>S:  How do they spend most of his time sitting?</td>
</tr>
<tr>
<td>T:  You’re very close. The question would be, “How does spinner’s mate spent most of his time?” Now you ask it.</td>
</tr>
<tr>
<td>S:  How does spinner’s mate spend most of his time?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>S:  How does the pressure from below push the mass of hot rock against the opening? Is that it?</td>
</tr>
<tr>
<td>T:  Not quite. Start your question with “What happens when?”</td>
</tr>
<tr>
<td>S:  What happens when the pressure from below pushes the mass of rock against the opening?</td>
</tr>
<tr>
<td>T:  Good for you! Good job.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>S:  What is the most interesting of the insect-eating plants, and where do the plants live at?</td>
</tr>
<tr>
<td>T:  Two excellent questions! They are both clear and important questions. Ask us one at a time now.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Day 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>S:  Why do scientists come to the south pole to study?</td>
</tr>
<tr>
<td>T:  Excellent question. This is what this paragraph is all about.</td>
</tr>
</tbody>
</table>
of 2–47, four students scoring at or below the 25th percentile. Although performance on these probes was variable, the total mean change in percentile rankings (combined across settings and phases) were Student 1, 47; Student 2, –.5; Student 3, 26; Student 4, 35.5; Student 5, 40.6; and Student 6, 36. Excluding Student 2, at the conclusion of the study, the range of subjects’ mean percentile ranks was 49–76. All students, except Student 2, demonstrated considerable generalization to the classroom setting. Student 2 was also the only student who did not reach criterion during the intervention.

Transfer tests were conducted in a pre- and posttest format. It would be impossible to go into all the details of the transfer probes here (see Palincsar & Brown, 1983, for details). Briefly, three of the four tests showed a significant improvement: writing summaries, designing questions to be asked on a test, and error detection using the Harris et al. (1981) procedure. The students did not improve on the Brown and Smiley (1977) task of rating narratives for variations in importance, although they did improve in their ability to select important elements in their summary writing.

To give only the flavor of the transfer results, we consider one test, the question-prediction task. The ability to generate important and clear questions was a skill that received considerable focus during training. A transfer measure was included to assess the accuracy with which students could identify and construct “teacher-like” questions. The students were given 4 randomly assigned passages, 2 prior and 2 following the study. Students were asked to predict and write 10 questions a classroom teacher might ask if testing the student’s knowledge of the passage.

The pre- and posttest scores are shown in Figure 2 for the students of Studies 2 and 3 and also for an untreated control group. The comparison group on the right of the figure represents the level set by average seventh-grade readers on this task. Training brought the level of performance up to that set by the normal comparison group. The graph is designed to illustrate where the improvement was found. Trained students improved in the overall quality of their questions, in the match between their questions and those actually generated by teachers, in their ability to paraphrase rather than lift questions directly from the text, and in their ability to concentrate on the main ideas.

In summary, the main findings of Study 2 were that students diagnosed as experiencing particular problems with reading comprehension improved considerably as a result of taking part in the reciprocal-teaching sessions. All students reached asymptote within 15 days, and for 5 of the 6 the level was at 70–80% correct, comparable to accuracy attained by 13 good comprehenders who acted as control subjects. Only Student 2 failed to reach the normal level, but she did improve from 12 to 50% and maintained that level well. Indeed, all of the students maintained their asymptotic level for at least 8 weeks.

In addition to this dramatic increase on the daily comprehension measures, the students improved their percentile ranking in the classroom, gaining an average of 37 percentile points. The quantitative improvement in the ability to
answer comprehension questions on texts read in a variety of settings was accompanied by a qualitative improvement in the students’ dialogues. Main-idea statements and summaries came to predominate, and unclear, incomplete, or detail responses dropped out.

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 3.** Given the success of Studies 1 and 2, another replication was attempted, but this time the teacher would be a “real” teacher, not an investigator, and the instruction would take place in naturally occurring groups within the school setting. In Study 3, four groups of students were considered, two classroom reading groups for the poorest readers and two reading groups that met regularly in a resource room. The group size ranged from 4 to 7 students. In all other respects the study was a replica of Study 2.

The teachers received three training sessions. In the first, they were introduced to the rationale behind the reciprocal-teaching intervention and were shown the results of Study 1. They also viewed a videotape of the investigator employing the technique with a group of students.

In the second training session, the teacher and the investigator practiced the procedures privately with the investigator modeling both the teacher’s role and behaviors 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, the teacher and the investigator met with a group of seventh graders who were not taking part in the study and practiced the procedure. The investigator modeled how the procedure would be introduced to the students and modeled the four main activities and the process of feedback. The teachers then assumed responsibility for the group, and as the practice session transpired, the teacher and investigator discussed the proceedings with each other. In addition, the teachers were left with several pages of directions regarding the introduction and daily format of the training sessions. The investigator 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 encountered. 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. All reciprocal-teaching sessions were tape-recorded.

The four groups of subjects were subjected to different amounts of baseline (4–10 days); otherwise, they were treated identically. Individually, the students performed in a manner similar to that found in Study 2 (see Palincsar & Brown, 1983, for full details). All of the subjects in Groups 1–3 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, 2 of the 4 students were performing quite well on the first day. The resultant group in some sense consisted of three models, the teacher and the 2 good students, and 2 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. Such findings if replicated could have important implications for decisions concerning the composition of the “optimal” reading group. All improvements were maintained over both short- and long-term follow-up sessions.

Quality of dialogue. An improvement in quality of dialogue was found as in Studies 1 and 2 but was less dramatic in Study 3. In the group settings, the teachers decided to call upon the “better students” in the initial sessions and then gradually to introduce the poorer students into the dialogue as they felt they could handle the responsibility, a natural procedure for experienced teachers. This resulted in a level of student responses that was higher initially and did not improve as dramatically over sessions. The trend was still the same, however, with incomplete or unclear questions decreasing from 20% to 4% and main-idea questions increasing from 57% to 70% across the sessions. Similarly, main-idea summaries increased from 68% to 85% of the total produced by the groups.
Transfer Tests
The same pattern of transfer performance occurred in Studies 2 and 3. Reliable improvements were found on three of the four tests: writing summaries, predicting questions, and error detecting. Again training brought the level of performance up to that set by normal seventh-grade readers.

The Palincsar and Brown series of studies can be regarded as successful for six main reasons:

1. The effect was large and reliable; of the 10 subjects included in Studies 1 and 2, 9 improved to the level set by good comprehenders, and all of the subjects in Study 3 met this level.

2. The effect was durable; maintenance probes showed no drop in the level of performance for up to an 8-week period (Studies 2 and 3). Although there was a decline after 6 months (levels dropping from 70–80% to 50–60%), only one session with the reciprocal-teaching procedure was sufficient to raise performance back to the short-term maintenance level (Study 1).

3. The effect generalized to the classroom setting; of the 10 students taking part in Studies 1 and 2, 9 showed a clear pattern of improvement, averaging a 36 percentile-rank increase, thus bringing them up to at least the average level for their age-mates. Given the difficulty reported in obtaining generalization of trained skills across setting (Brown & Campione, 1978; Meichenbaum & Asarnow, 1978), this is an impressive finding.

4. Training resulted in reliable transfer to dissimilar 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; Brown, Campione, & Day, 1981).

5. Sizable improvements in standardized comprehension scores were recorded for the majority of subjects.

6. The intervention was no less successful in natural group settings conducted by teachers than it was in the laboratory when conducted by the experimenter.

Training Studies and the Problem of Multiple Determinants
Let us consider some possible reasons for the success of the Palincsar and Brown studies when so many other attempts have failed to find durability, generalization, and transfer of the effects of training. First, the training was extensive. Second, the activities trained were well specified theoretically, and well established empirically as particularly problematic for poor readers. Third, the training was specifically tailored to the needs of these particular students, good decoders but passive comprehenders. Fourth, the skills themselves could reasonably be expected to be trans-situational. Such ubiquitous activities of self-review and self-interrogation are pertinent in a wide variety of knowledge-acquisition tasks.
In addition, a great deal of attention was paid to “metacognitive” variables (Baker & Brown, 1984a): The subjects were fully informed about the reasons why these activities were important; the subjects were given explicit information concerning the generality of the activities and their range of utility; the subjects were trained in self-regulatory activities including the checking and monitoring of their own comprehension; and the skills themselves were general comprehension-monitoring activities applicable in a wide variety of reading/studying tasks.

The reciprocal-teaching mode itself could be responsible for the improvement. The interactive format permits extensive modeling of the target activities in a reasonably natural setting. It also forces the students to participate at whatever level they can so that the teacher can evaluate current states and provide appropriate feedback and assistance (refer to Table 3).

Listing all the good points about the Palincsar and Brown studies leads us to the obvious problem of interpretation. The studies are multiply confounded, and this is true to some extent of all the successful cognitive-training studies to date (Chipman, Segal, & Glaser, 1985). For example, would a single activity rather than the package of paraphrasing, questioning, predicting, and clarifying have been successful? Component analyses studies currently underway in our laboratory suggest that whereas all of the activities engaged in individually result in improvement, the summary component is the most powerful. The combined package, however, is the most effective intervention.

Similarly, the addition of the metacognitive setting variables may or may not be essential and such variables permit of degrees. For example, in a Ph.D. thesis conducted in our laboratory, Day (1980) trained junior college students to apply basic rules of summarization and to check that they were using the rules appropriately. The subjects were remedial students who, although of normal reading ability, were diagnosed as having writing problems. There were three main instructional conditions that varied in how explicit the training was:

1. **Self-management**: The students were given general encouragement to write a good summary, to capture the main ideas, to dispense with trivia and all unnecessary words—but they were not told rules for achieving this end.

2. **Rules**: The students were given explicit instructions and modeling in the use of the rules of deletion, selection, invention, etc.

3. **Control of the rules**: The third and most explicit training condition involved training in the rules and additional explicit training in the control of these rules; i.e., the students were shown how to check that they had a topic sentence for each paragraph, how to check that all redundancies had been deleted, all trivia erased, etc., and how to check that any lists of items had been replaced with superordinates, etc.

An example of the results is shown in Figure 3, where the data from one of the rules, selection, are shown. The degree of posttest improvement was significantly related to the explicitness of training. Merely telling students to stay on task,
be economical, concentrate on main ideas, i.e., the self-management condition, produced significantly less improvement than did direct instruction in using the specific rules, which in turn was less successful than a combined package that involved both practice using the task-appropriate rules and direct instruction in monitoring and overseeing their application. In this context it should be noted that in the Palincsar and Brown studies, the students not only received modeling of the appropriate comprehension-fostering activities, they were also explicitly and repeatedly directed to use these activities while reading on their own.

There is growing evidence that the most successful cognitive-skills training packages will include three components: (1) skills training, practice in the use of the task-appropriate cognitive skills; (2) self-control training, direct instruction in how to orchestrate, oversee, and monitor the effective use of the skills; and (3) awareness training, information concerning the reasons why such strategy use improves performance and detailed instruction in when and where the strategies should be used. For practical reasons, interventions should include all these factors (Brown & Palincsar, 1982), but for theoretical reasons, we need to conduct component analyses of the separate effect of all three forms of metacognitive settings (Brown, Bransford, Ferrara, & Campione, 1983).

The reciprocal-teaching package is also multiply confounded. Would modeling alone, feedback alone, or just explicit instruction be as effective? Such component analyses studies are currently underway in our laboratory. Preliminary
evidence again favors the combined package (see also Bird, 1980), but more data are needed.

From a practical standpoint, the results of the required component analyses would be helpful in permitting the streamlining of the training packages into efficient and economical units. From a theoretical perspective, we need considerable further research before we can attribute the success of the intervention appropriately. Of course, it could be that multiply confounded interventions are needed because successful reading comprehension is a multiply determined outcome; i.e., effective comprehension rests on the interaction of a number of “separate” activities. Given the typically limited outcome of restricted cognitive-skills training studies (Brown, Campione, & Day, 1981), we advocate the procedure of first obtaining an educationally relevant, sizable, durable, and generalized effect of training and then conducting the necessary investigations to determine the subcomponents that are primarily responsible for the improvement.

The Theoretical and Practical Status of Training Studies

Until the mid-1970s, the prognosis of worthwhile educational gains from cognitive-skills training studies was poor. Although some success had been achieved in obtaining improvement on a particular skill in isolation, this improvement was often slight and fleeting, and there was very little evidence of transfer. Maintenance over time, generalization across settings, and transfer within conceptual domains were rarely found. The more difficulties the learner experienced initially, the more fleeting and bounded were the effects of training (Brown & Campione, 1978, 1981; Meichenbaum & Asarnow, 1978). But the picture has changed in the last few years; the success of the Palincsar and Brown studies is not an isolated phenomenon (Chipman, Segal, & Glaser, 1985). The current outlook is quite optimistic. From a practical point of view, it is clear that we can train instructionally relevant cognitive skills even with subjects who would be regarded as recalcitrant. This training can be carried out under the pressure of normal classroom settings, and it does result in worthwhile and reliable improvements in the Palincsar and Brown studies, bringing students from the very bottom of the distribution of their age peers to the average set by their normal reading classmates. The necessary research needed now consists of extensions across skills and settings and of “clean-up” operations that would permit us to test the limits of these exciting findings and streamline our instructional packages. Cognitive skills can be trained, and such training can be durable and generalizable.

From a theoretical standpoint, training studies are not just exercises in cognitive engineering with immediate applicability to school settings. They are also direct tests of theory involving degrees of experimental manipulation and control in an area where a great deal of data consist of simple one-shot developmental demonstrations (see chapter 9 of Learning and Comprehension of Text, 1984, for a detailed discussion of this argument). A great deal of development research is correlational in nature, and there are problems with interpreting such results. To give an example from our own work, in many studies we consider the performance of
students who do or do not spontaneously adopt an appropriate text-processing strategy, and this is often the major variable carrying a developmental trend. For example, 5th and 7th graders who make adequate rough drafts when paraphrasing (Brown, Day, & Jones, 1983) or spontaneously underline or take notes of important text elements and so on (Brown & Smiley, 1978) perform as well as the majority of 12th graders, whereas 12th graders who fail to employ these activities perform more like 5th graders. This pattern suggests that it is the strategy that leads to efficiency, and developmental trends showing improvements with age are created by the increased proportion of strategic subjects. This is a reasonable interpretation, but as the data are primarily correlational, the interpretation is not that simple. 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 partialling out ability factors such as IQ or reading scores does not totally bypass this problem.

The training study is then an important tool for providing convergent evidence of the importance of the strategy under consideration. First, the theorist speculates about the underlying processes involved in reading comprehension. Next is the correlational step; students who read well are also found to perform well on the identified underlying processes, whereas poor readers experience particular difficulty on just these activities (Armbruster, Echols, & Brown, 1982; Baker & Brown, 1984a, 1984b; Brown & Palincsar, 1982). Finally, students who are not using the strategy are given training designed to induce the use of processes theoretically specified as key activities underlying efficiency. Others are not. If the theory is correct (and training adequate) and these are the underlying effective processes, trained students’ performances should become more like those of spontaneous users. There are nontrivial problems with interpreting the outcomes of training studies (Brown & Campione, 1978, 1981), but they do provide an important manipulative tool to aid theory development. Thus 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 mechanism of reading comprehension.

**QUESTIONS FOR REFLECTION**

1. What processes did the authors of reciprocal teaching go through to create the strategy?
2. Why is it difficult to isolate one specific strategy or instructional tool that will help poor readers become better at comprehension?
3. How did the authors confirm the efficacy of reciprocal teaching?
4. What might be the challenges of replicating the elements of the authors’ training studies in your own classroom?
ACKNOWLEDGMENTS
The preparation of this manuscript was supported in part by Grants HD06864 and HD05951 from the National Institute of Child Health and Human Development and in part by NIE-C-400-76-0116 from the National Institute of Education.

NOTES
This chapter is reprinted from Learning and Comprehension of Text (pp. 255–286), by H. Mandl, N.L. Stein, & T. Trabasso, Eds.). Copyright © 1984 by Lawrence Erlbaum Associates. Reprinted with permission.

1 The students were of low-normal intelligence (mean IQ 84) and low socioeconomic status. Their decoding was judged adequate as they could read grade-appropriate texts at a rate of 80–100 wpm with no more than 2 errors per minute. Their standardized reading comprehension scores averaged 3 years delayed.

2 In Study 1, another group of students received a second intervention, locating information (see Palincsar & Brown, 1983, for details), where they were trained to answer comprehension questions by using the text intelligently. These students did improve from their starting level of 15% to approximately 50%, but they never reached the level of the reciprocal-teaching group, and they failed to maintain this level over time.

3 In Studies 2 and 3, there were treated and untreated control groups consisting of students matched with the experimental subjects for decoding and comprehension scores, as well as IQ, standardized tests, and class placement. These students demonstrated no significant change on their performance on the baseline, maintenance, and follow-up stages of the study. Neither did they improve their performance on any of the tests of generalization (to the classroom) or transfer (across laboratory tasks). For full details on the control groups included in Studies 1 to 3, see Palincsar and Brown (1983).

REFERENCES
References


**Postscript on Reciprocal Teaching**

*Annemarie Sullivan Palincsar*

I am thankful to the editors for this opportunity to include a sequel to the chapter “Instructing Comprehension-Fostering Activities in Interactive Learning Situations.” Whereas the original chapter was comprehensive in its attention to comprehension instruction, this postscript focuses on reciprocal teaching (RT), specifically on the evolution of the intervention itself, as well as the evolution of research on RT. Finally, I speak to the evolution of the field with respect to comprehension instruction, as informed by research on RT.

There are two transformations in the design and use of RT that seem noteworthy. One was the introduction of thematically related passages as the grist for the dialogues. The initial studies of RT employed grade-appropriate texts that address a broad array of (fairly random) unrelated topics; the limitation of this choice did not promote the use of the dialogues for knowledge building over time. In contrast, in a line of inquiry conducted with first-grade students in which RT was used to teach listening comprehension, we wrote texts that presented simple science concepts related to animal survival: protection from elements, natural pest control, adaptation and extinction, camouflage, and mimicry. For example, the theme protection against elements included passages about porcupines, turtles, and armadillos. Furthermore, the teacher was urged to first determine whether the students would spontaneously focus on the content of the theme.
in their discussion of the passage. If the students failed to do so, the teacher was encouraged to make the theme explicit (Palincsar & Brown, 1988; Palincsar, Stevens, & Gavelek, 1989).

The results of the listening comprehension assessments that we administered were that the students in the RT group scored 70% correct after 20 days of discussions, whereas the students in the comparison condition, who listened to each story but did not have the benefit of discussion, scored 40% correct on these measures. Furthermore, the students in the experimental group were significantly more successful at identifying the gist of passages and recognizing and applying analogical information in the test passages to novel passages (Palincsar, Brown, & Campione, 1993).

A second curricular modification was the inclusion of RT as one participation structure within Brown and Campione's (1994, 1996) Fostering a Community of Learners research. In these classrooms, students were prepared to use RT dialogues in the same fashion that students in earlier research were prepared, but RT was then incorporated in a much larger and more ambitious curriculum: Student groups would become experts regarding particular topics (e.g., interdependence in an ecosystem), and the students used RT dialogues in the course of conducting their research and sharing their research with others. Both of these curricular adaptations are significant with respect to what they communicate about the ultimate goal of RT dialogues, a goal that will be revisited shortly in this postscript.

The earliest characterizations of RT indicated that there were three essential components to understanding and replicating RT: the four strategies that were complementary and designed to accomplish different work in the context of constructing the meaning of text, the teacher–student/student–student dialogue that was key to the collaborative construction of meaning, and the roles of the teacher in the dialogues. These roles were characterized as multifaceted and included diagnosing difficulties arising in understanding the text, and adjusting support to enact the strategies as needed. Ideally, as teachers model these roles, students will begin to assume greater responsibility for assuming these roles themselves. Studies that followed the initial studies of RT were designed to investigate the interaction of these three components. For example, Palincsar (1986) reported analyses of teacher–student dialogues, examining groups of first graders and their teachers, all of whom were using RT with the same texts but with more and less success. Features characterizing the teachers’ practices in the more successful groups included requesting that students elaborate on their ideas, restoring direction to the discussion when it began to meander, and reworking students’ contributions so they were integrated into the discussion.

Perhaps as a function of how deceptively simple RT looks on the outside, certainly as a function of not having sufficient supports for dissemination, and perhaps as a function of being a significant departure from typical classroom practice, as RT was disseminated, it fell prey to what Haertel (1986) has called a series of “lethal mutations” (p. 52). Perhaps because they are the easiest part of the instruction to get a handle on—certainly in contrast to the discourse in which

686 Brown, Palincsar, and Armbruster
they were embedded—the four strategies appear to have taken on a life of their own; they appear in standards, in reading texts, in worksheets, and in language arts series. A quick Internet search reveals many examples of videos in which groups purport to be engaged in RT dialogues, but there is nothing dialogic about the instruction. Typically, what is demonstrated is practice in the use of the four strategies, with little evidence that the group is using the strategies to construct a coherent representation of the ideas in the text.

Spurred by this “lethal mutations” phenomenon and seeking a more powerful way to accurately communicate the characteristics of RT, Palinscar and Spiro collaborated on the design of a Web-based hypermedia environment, which Spiro calls an experience acceleration support environment, or EASE system (Spiro & Jehng, 1990). We named our EASE system Teaching Text, Making Meaning (TTMM; see edr1.educ.msu.edu/CompStrat/login.asp, username: demo, password: demo). TTMM was developed through an iterative coding process in which we analyzed 35 hours of reading comprehension instruction dialogues with upper elementary students; the 35 hours were converted to 188 short video clips grouped into categories, which we refer to as themes. The 10 themes (e.g., building a learning community, modeling expert reading, using text characteristics) are further divided into specific topics. For example, the using text characteristics theme includes clips about genre, text structure, and text features (e.g., illustrations). The 10 themes and their associated topics were selected because they address the issues that teachers face as they plan, enact, and evaluate text-based discussions. For example, the using text characteristics theme illustrates how a teacher might make use of an illustration in the text to support students’ understanding of the content. Linked to a number of the instructional clips are related interviews in which the teachers reflect on the instructional moment featured in each clip, such as evaluating their practice, describing tensions in the instruction, or identifying alternative decisions entertained in the moment.

The navigability of an EASE system raises interesting questions about communicating about teaching, specifically about the teaching of reading comprehension. The video clips are not primarily organized around a particular instructional approach; rather, they are organized around moments of teaching and learning. Assembling this environment challenged us to think about useful grain sizes for parsing the work of teaching and the development of competence as text comprehenders. The hypermedia tool gives teachers the ability to navigate the system as a function of their own needs and interests; furthermore, they can use the system to engage in inquiry, essentially bringing their own question(s) to the data, which, in this case, are the video clips. Analyses of teachers engaged in use of the system reveals evidence of increased sensitivity to context and a greater appreciation for the nuanced nature of the application of practices such as scaffolding. For example, teachers were observed shifting over time from searching for a singular and prescriptive definition of scaffolding to a more situation-sensitive analysis of the forms and purposes for scaffolding (Palincsar et al., 2007).
In addition to the research conducted by its developers, RT has a broad research base that has been built by others (Rosenshine & Meister, 1994). Although originally designed for use with elementary students scoring at or below the 35th percentile on standardized reading measures, RT has been implemented and found effective in a variety of contexts and for a variety of readers. The strategy has been effective in teaching students with mild disabilities in resource (Marston, 1995) and inclusive settings (Lederer, 2000), deaf and hard-of-hearing students (Al-Hilawani, 2003), high school students (Alfassi, 1998; Westera & Moore, 1995), bilingual students in the United States (Padrón, 1992), and international students who are learning English as a second language (Fung, Wilkinson, & Moore, 2003).

Of course, RT is only one of many instructional approaches that have been advanced by the literacy community to improve reading comprehension. It is, therefore, disappointing that descriptive studies reveal that comprehension instruction continues to be enacted infrequently (Connor, Morrison, & Petrella, 2004; Taylor, Pearson, Clark, & Walpole, 2000) and inadequately (Dewitz, Jones, & Leahy, 2009). Furthermore, despite decades of research testifying to the effectiveness of teaching readers to be self-regulating or strategic in their reading of text, strategy instruction itself has come under attack by members of the literacy community (e.g., McKeown, Beck, & Blake, 2009; Wilkinson & Son, 2011). As the history of RT testifies, strategy instruction becomes vulnerable when the rigid or isolated teaching of strategies becomes the goal of instruction, rather than the flexible and opportunistic application of strategies to sense making with text. The cumulative research on RT suggests that features of powerful comprehension instruction includes the use of related texts that allow students to experience the process of deepening their understanding across texts; the teaching of strategies as a repertoire of thinking tools that should be used in opportunistic ways, as determined by the demands of the text and the goals of the reader; and the explicit teaching of strategies in the service of advancing knowledge building.

REFERENCES


