Effect of Think-Aloud Instruction on Elementary Students' Comprehension Monitoring Abilities
James F. Baumann, Nancy Seifert-Kessell and Leah A. Jones
*Journal of Literacy Research* 1992 24: 143
DOI: 10.1080/10862969209547770

The online version of this article can be found at:
http://jlr.sagepub.com/content/24/2/143

Published by:
*SAGE*
http://www.sagepublications.com

On behalf of:
Literary Research Association

Additional services and information for *Journal of Literacy Research* can be found at:

**Email Alerts:** http://jlr.sagepub.com/cgi/alerts

**Subscriptions:** http://jlr.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

**Citations:** http://jlr.sagepub.com/content/24/2/143.refs.html

>> Version of Record - Jun 1, 1992

What is This?
EFFECT OF THINK-ALOUD INSTRUCTION ON ELEMENTARY STUDENTS' COMPREHENSION MONITORING ABILITIES

James F. Baumann  
*University of Georgia*

Nancy Seifert-Kessell  
*Cleveland State University*

Leah A. Jones  
*Chandler, Arizona*

**ABSTRACT**

This study investigated the effectiveness of explicit instruction in think aloud as a means to promote elementary students' comprehension monitoring abilities. Sixty-six fourth-grade students were randomly assigned to one of three experimental groups: (a) a Think-Aloud (TA) group, in which students were taught various comprehension monitoring strategies for reading stories (e.g., self-questioning, prediction, retelling, rereading) through the medium of thinking aloud; (b) a Directed Reading-Thinking Activity (DRTA) group, in which students were taught a predict-verify strategy for reading and responding to stories; or (c) a Directed Reading Activity (DRA) group, an instructed control, in which students engaged in a noninteractive, guided reading of stories. The primary quantitative analyses involved two planned orthogonal contrasts—effect of instruction (TA + DRTA vs. 2 × DRA) and intensity of instruction (TA vs. DRTA)—for three whole-sample dependent measures: (a) an error detection test, (b) a comprehension monitoring questionnaire, and (c) a modified cloze test. Results of effect of instruction contrasts revealed that TA and DRTA students were more skillful at comprehension monitoring than DRA students (TA + DRTA > DRA for all three measures). Results of intensity of instruction contrasts indicated that although TA-trained students had greater awareness of comprehension monitoring abilities (TA > DRTA for the questionnaire), DRTA students' performance equaled (TA = DRTA for the cloze test) or exceeded (TA > DRTA for the error detection test) that of the TA students. Qualitative data from student interviews, however, revealed that TA students both reported and demonstrated using a greater depth
and breadth of comprehension monitoring abilities than either DRTA or DRA students. It was concluded that both TA and DRTA strategies are effective for enhancing elementary students' comprehension monitoring abilities but that additional research is needed to determine the relative effectiveness of TA and DRTA approaches.

This article reports the results of an instructional experiment, the purpose of which was to evaluate the effectiveness of explicit instruction in thinking aloud as a means to enhance young students' comprehension monitoring abilities. The empirical background for the experiment draws from research on comprehension monitoring and the think-aloud procedure.

Baker and Brown (1984a) stated that comprehension monitoring is that dimension of metacognition which involves a learner's use of self-regulation in receptive language processing (listening and reading). Although definitions of comprehension monitoring during reading vary (cf. August, Flavell, & Clift, 1984; Baker & Brown, 1984b; Brown, 1975; Garner, 1987; Paris, Lipson, & Wixson, 1983; Pitts, 1983; Weinstein & Rogers, 1985), most include two kinds of metacognitive knowledge: (a) a reader's awareness of whether or not comprehension is occurring, and (b) a reader's conscious application of one or more strategies to correct comprehension difficulties (Baker, 1979; Baker & Brown, 1984a; Brown, 1980; for a synthesis, see Wagoner, 1983). Simply stated, comprehension monitoring "concerns the student's ability both to evaluate his or her ongoing comprehension processes while reading through a text, and to take some sort of remedial action when these processes bog down" (Collins & Smith, 1982, p. 174).

Research indicates that comprehension monitoring abilities discriminate successful readers from less successful ones (e.g., August et al., 1984; Paris & Myers, 1981; see Brown, Armbuster, & Baker, 1986, for a review), and studies demonstrate that learners can be taught a variety of metacognitive strategies that enhance understanding (see Paris, Wasik, & Van der Westhuizen, 1988, for a review). For example, students have been taught specific strategies such as inference training (Dewitz, Carr, & Patberg, 1987), mental imagery instruction (Gambrell & Bales, 1986), among others (Babbs, 1984; Miller, 1985, 1987; Miller, Giovenco, & Rentiers, 1987; Nolte & Singer, 1985; Raphael & Pearson, 1985; Raphael & Wonnacott, 1985; Reis & Spekman, 1983). Other experiments have involved instruction in multiple metacognitive and comprehension monitoring strategies (Lysynchuk, Pressley, & Vye, 1990; Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987; Paris, Cross, & Lipson, 1984; Paris & Jacobs, 1984; Paris & Oka, 1986; Paris, Saarnio, & Cross, 1986; Schmitt, 1988; Tregaskes & Daines, 1989).

One technique that has been used to evaluate comprehension monitoring abilities is the oral think-aloud procedure. Think alouds involve the overt, verbal expression of the normally covert mental processes readers engage in when constructing meaning from a text (Afflerbach & Johnston, 1986; Ericsson & Simon, 1980, 1984;
Garner, 1987). Think alouds have been used in descriptive research on writing processes (Afflerbach, Bass, Hoo, Smith, Weiss, & Williams, 1988; Hayes & Flower, 1980, 1983), in descriptive research on reading processes (Alvermann, 1984; Bereiter & Bird, 1985, Study 1; Garner & Alexander, 1982; Hare & Smith, 1982; Olshavsky, 1976–1977; Randall, Fairbanks, & Kennedy, 1986; Scardamalia & Bereiter, 1984), and they have been recommended as a comprehension monitoring assessment tool (Baumann, 1988; Brown & Lytle, 1988; Wade, 1990).

Several writers have proposed teaching students to think aloud as a means to enhance comprehension monitoring abilities (Alvermann, 1984; Davey, 1983; Nist & Kirby, 1986). However, intervention research on think alouds has been limited to a few studies. Thurmond (1986) used think alouds to teach verbal reasoning and vocabulary skills to 22 eleventh-grade minority high school students who were enrolled in an 8-week summer honors course offered by a medical school as a means to attract minorities into health care professions. Thurmond reported significant pretest-posttest differences on the Scholastic Aptitude Test and the Nelson-Denny Reading Test, although lack of methodological detail precludes a more thorough analysis of this experiment.

Bereiter and Bird (1985, Study 1) analyzed protocols of adults’ think alouds during reading and identified four strategies: (a) restatement (paraphrasing and summarizing), (b) backtracking (looking back and rereading), (c) demanding relationships (self questioning and inferring unstated information), and (d) problem formulation (hypothesizing and predicting).

In Study 2, Bereiter and Bird (1985) sought to determine the “teachability of the reading strategies identified in Study 1 and in particular to assay the prospects for use of thinking aloud as an instructional vehicle” (p. 140). Seventh- and eighth-grade students were assigned to one of four groups: (a) modeling plus explanation—students were taught the four think-aloud/reading strategies through verbal explanation and modeling by the teacher; (b) modeling only—identical to the preceding treatment except the teacher provided no verbal explanation; (c) exercise—students were provided oral and written exercises requiring them to engage in the four strategies, but they were provided no explanation or modeling of those abilities; or (d) uninstructed control—students participated in pre- and posttesting only.

Intervention students engaged in nine 40-minute sessions over 3 weeks. Following intervention, two posttests were administered to all subjects. The Nelson-Denny Reading Test was administered to assess silent reading comprehension. Strategy use (restatement, backtracking, etc.) and oral comprehension were assessed by having the students think aloud while reading six passages from the comprehension portion of the Metropolitan Achievement Test and then respond orally to the comprehension questions.

Results revealed that comprehension scores on both the Nelson-Denny Reading Test and the Metropolitan Achievement Test were significantly higher for the
modeling-plus-explanation group compared to the other three groups, which did not differ significantly from one another. Analyses of the think-aloud protocols for the Metropolitan Achievement Test passages revealed that students in the modeling-plus-explanation group produced more think alouds of any kind (target strategies plus others) than control group subjects. When only the four target strategies were examined, results revealed that the modeling-plus-explanation group surpassed the modeling-only group for restatement, backtracking, and problem formulation; there were no differences between modeling-only subjects and controls. From these findings, Bereiter and Bird (1985) drew the following conclusions:

The present study offers strong evidence that students will not readily acquire cognitive strategies simply by imitating models and that they also need direct, explicit instruction in the strategies . . . . What is clear from the present study is that modeling and direct instruction combine very naturally in the teaching of cognitive strategies. Direct instruction appears to be important for getting students to attend to what is relevant in cognitive modeling. (p. 153)

The study reported here was conducted to extend the work of Bereiter and Bird (1985), specifically to investigate the efficacy of explicit instruction in thinking aloud as a means to enhance young children’s comprehension monitoring during reading. Although there is ample evidence that elementary age students can be taught various comprehension monitoring abilities, no think-aloud intervention research has been done with students below Grade 7; therefore, younger students (fourth graders) were chosen as subjects. Because of Bereiter and Bird’s success with direct instruction in thinking aloud that included significant teacher modeling, an explicit instruction approach that also included teacher modeling (Baumann & Schmitt, 1986) was selected for the primary intervention in the experiment, the Think-Aloud (TA) group. This group received training in several comprehension monitoring and correction strategies through the instructional vehicle of think aloud.

The Directed Reading-Thinking Activity (DRTA) (Stauffer, 1969, 1976) was selected as the comparison intervention strategy because DRTA involves significant amounts of prediction, a strategy central to most descriptions of comprehension monitoring (e.g., Collins & Smith, 1982). Further, prediction is a component of several successful programs for teaching metacognitive strategies (e.g., Palincsar & Brown, 1984; Palincsar, Brown, & Martin, 1987; Paris et al., 1984). Thus, the efficacy and efficiency of the intensive TA instruction could be evaluated in comparison to a relatively simple strategy that had the potential to also induce comprehension monitoring behaviors.

An instructed control group was included in the experiment to minimize the likelihood of a Hawthorne effect, a serious problem in comprehension strategy research (Lysynchuk, Pressley, d’Ailly, Smith, & Cake, 1989). Students in this group read the same stories as the TA and DRTA groups according to the Directed Reading Activity (DRA) (Tierney, Readence, & Dishner, 1990). The DRA was an
appropriate comparison group because it did not involve any explicit or implicit instruction in metacognitive or comprehension monitoring behaviors.

In summary, the purpose of this study was to evaluate the effectiveness of explicit instruction in a think-aloud procedure as a means to promote young students’ comprehension monitoring abilities relative to the DRTA and DRA procedures. Specifically, two research questions were posed: (a) Does TA training or the use of the DRTA enhance students’ ability to monitor their reading comprehension when compared to the DRA? and (b) Is there any difference between TA instruction and DRTA instruction in promoting students’ comprehension monitoring ability?

METHOD

Subjects and Design

Fourth-grade students from an elementary school in a rural midwestern community participated in the study. The school had adopted a team teaching plan, and three fourth-grade teachers shared an open classroom area. The teachers grouped the students for reading instruction across the three classes according to reading ability. One teacher each worked with a high, middle, and low reading group. All groups used basal reading materials and were formed at the beginning of the school year on the basis of standardized achievement test data, prior performance in the basal reading program, and teacher judgment.

Sixty-eight of the 72 students in the fourth-grade team participated in the study; the four nonparticipating students were mainstreamed special education students for whom the instructional materials were considered to be too frustrating to read. The 68 students were assigned randomly to one of the three experimental groups (TA = 23 students; DRTA = 23 students; DRA = 22 students). Because of illness and transfer to another school, incomplete data were obtained for one subject each from the TA and DRTA groups. Thus, the final sample consisted of 66 students, 32 girls and 34 boys.

A pretest-posttest control group design (Campbell & Stanley, 1963) was employed in which the independent variable was treatment (TA, DRTA, DRA) and the dependent variables were quantitative scores and qualitative results from four posttests designed to evaluate students’ think aloud and comprehension monitoring abilities. Posttests 1–3 were group-administered, quantitative measures obtained for all students in the sample; Posttest 4 involved interview data that were obtained for 12 subjects, 4 from each treatment. To statistically account for students’ preexperimental differences in comprehension monitoring abilities, two pretests were constructed, administered, and used as covariates in data analyses for Posttests 1–3.
Materials

Pretest 1. The first pretest involved an error detection task designed to evaluate students' preintervention ability to monitor their comprehension. Students were presented with a basal reader story within which 16 sentences were intruded. (All basal materials in the experiment were taken from the two third-grade books of the Ginn Reading Program, Clymer, Indrisano, Johnson, Pearson, & Venezky, 1987.) The intruded sentences were semantically consistent with the selection at a lexical level in that they included many of the same characters, events, and settings presented in the story. However, they were semantically inconsistent with the selection at a more global level. For example, the sentence, "You're a nice dog, Boots" was intruded in Pretest 1, a selection about a young girl's attempt to find homes for a litter of kittens recently born by her pet cat named Boots. Clues for 8 of the 16 inconsistencies came before the intrusions; clues followed the intrusions for the other eight inconsistencies.

Students were provided the following directions which were printed on the pretest and read aloud by the experimenter:

Read the story carefully. Try to understand it as you read. There will be 16 sentences that do not belong in the story. Look for them as you read. Ask yourself, "What does not make sense in this story?" Underline the 16 sentences that do not make sense. Check your work when you are finished. If you cannot read a word, raise your hand, and we will help you.

Preceding Pretest 1, students were given a brief practice exercise that required them to identify obvious intruded sentences in two short texts. Pretest 1 was scored by calculating the number of correctly identified intruded sentences.

Although students were assigned randomly to treatments, analyses of variance (ANOVA) for both pretests were conducted to verify the general equivalence of the three groups at the onset of the experiment. The ANOVA for Pretest 1 failed to attain statistical significance, indicating no between-groups differences on this measure: $F(2, 63) = 1.13, p < .328$ (TA $M = 9.136, SD = 3.342$; DRTA $M = 9.727, SD = 2.694$; DRA $M = 10.506, SD = 2.972$). The reliability for Pretest 1, as calculated according to the Kuder-Richardson 21 (KR-21) formula, was $r = .62$.

Pretest 2. The second pretest queried students about the strategies they believed to be useful in promoting their understanding of stories. In this pretest, modeled after an instrument developed by Schmitt (1988, 1990), the students were presented with 15 multiple-choice items like the following:

When I read, it is a good idea to:
A. sound out words I don't know.
B. make some guesses about what will happen in the story.
C. make a list of all the details in the story.
D. look up new words in the dictionary.
Think-Aloud Instruction

Each item contained a response that was indicative of comprehension monitoring behaviors (option B above) such as self-questioning, prediction-verification, summarizing-retelling, paraphrasing, and rereading and reading on to clarify meaning. Other options focused on decoding, vocabulary, or noncomprehension-monitoring tasks (e.g., listing details from the story, looking for examples of cause/effect relations).

Students were given the following directions which were printed on the pretest and read aloud by the experimenter:

Think about what kinds of things you can do to help you understand a story better. Read each of the lists of four statements and decide which one of them would help you the most. There are no right answers. Circle the letter of the statement you choose.

To prevent poor decoding from interfering with students’ responses, the experimenter read all items aloud and provided the students time to respond to each item before proceeding to the next. The ANOVA on Pretest 2 indicated no between-group differences: $F(2, 63) = 0.11, p < .895$ (TA $M = 4.955, SD = 1.864$; DRTA $M = 5.091, SD = 1.998$; DRA $M = 5.273, SD = 2.763$). The KR-21 for Pretest 2 was $r = .36$.

Posttest 1: Error detection test. Posttest 1 was the same in form as Pretest 1: a basal story (though different from the story used for Pretest 1) with 16 intruded sentences lexically consistent at a local level but semantically inconsistent at a global level. Administration directions were identical to those for Pretest 1.

Because the students performed at a higher-than-expected level on Pretest 1 (61% of all intruded sentences were correctly identified), the experimenters were concerned about a potential post-intervention ceiling effect on this posttest, an occurrence which could mask group differences. To reduce the likelihood of a ceiling effect, the intruded sentences for Posttest 1 were written so their inconsistencies were more subtle than those included in Pretest 1, which explains the somewhat lower level of performance on Posttest 1 (51% of all intruded sentences correctly identified). For example, consider the intruded sentence (italicized) in the following paragraph from Posttest 1, in which the main character, Stuey, is experimenting with invisible ink which appears when heated by a light bulb:

As if by magic, clear blue words began to show. In a little while there was the urgent message: Report at once to Agent X. The password is Abracadabra. Stuey said, ‘Now I can put on a magic show!’

To identify this intrusion, students needed to recognize that the magic show was not referenced subsequently in the selection, which requires a level of comprehension monitoring more sophisticated than, for example, detecting that a cat named Boots is referred to later on in an intruded sentence as a dog (see earlier description of Pretest 1 item). The KR-21 for Posttest 1 was $r = .70$. 

Downloaded from jrt.sagepub.com at University of Kansas Libraries on March 29, 2014
Posttest 2: Comprehension monitoring questionnaire. Posttest 2, consisting of 18 items, was a slightly expanded version of Pretest 2. Three additional items were added to Pretest 2 and several others were revised so that there was better coverage and balance between the comprehension monitoring behaviors queried on the posttest and those actually taught to the TA group. Comprehension monitoring behaviors probed by the questionnaire included: (a) asking-answering questions (2 items), (b) predicting-verifying (3 items), (c) retelling and paraphrasing (4 items), (d) evaluating understanding (2 items), (e) rereading and reading on to clarify meaning (3 items), (f) using contextual information (textual and picture) to promote understanding (2 items), (g) thinking aloud (1 item), and (h) evaluating the source of information (1 item). Administration directions for Posttest 2 were identical to those used during pretesting. The KR-21 for Posttest 2 was \( r = .42 \).

Posttest 3: Degrees of Reading Power. The Degrees of Reading Power (DRP) (1986) is a standardized, commercially produced instrument that employs a modified cloze format (words are deleted from a selection and for each omission the student selects the most appropriate word from a list of five). According to the authors, DRP tests “are holistic measures of how well the messages within text are understood...” DRP tests focus measurement on determining how well students process or construct meaning from paragraphs as they read through a selection” (DRP Handbook, 1986, p. 1). Thus, the DRP, at one level, is a general reading comprehension assessment tool.

However, the DRP was selected as a posttest in this experiment because it taps aspects of comprehension monitoring. For example, for the DRP sentence, “The role of ______ is important,” the answer choices are wind, boats, weight, wires, and experience (DRP in Brief, 1985, p. 6), each of which is semantically and syntactically appropriate within the sentence. However, this sentence is embedded in a selection on bridges and river traffic, and the only semantically acceptable response within the context of the entire passage is boats. To select such a response, students must be metacognitively aware of global levels of coherence and comprehension and are likely to engage in behaviors such as rereading or withholding judgment and reading on. Employing this argument, several researchers (e.g., Paris, et al., 1984; Schmitt, 1988) have used cloze tests to evaluate students’ ability to monitor their comprehension. Further, Paris et al. (1984) found that cloze tests are consistently correlated with other measures of strategic reading (see Baker & Brown, 1984b, and Paris, 1991, for a more elaborate rationale), a finding corroborated in the present study, since the correlation between Posttest 1, the Error Detection Test, and Posttest 3, the DRP, was \( p < .001 \).

DRP form PA-8, a level intended for students in Grades 3–5, was administered. The number of correct items on the DRP (0–56) was the numerical value used in data analyses. The KR-21 for Posttest 3 was \( r = .85 \).

Posttest 4: Student interviews. In order to obtain more qualitative data on
students’ ability to think aloud and monitor their reading comprehension, 12 students (7 boys, 5 girls) were purposively selected (Manheim, 1977; Patton, 1980) from the larger sample for in-depth interviews. Specifically, four students from each treatment were selected such that: (a) they were verbal (as determined by experimenter and teacher judgment) and hence would be likely to communicate freely with the experimenters and produce responses of sufficient quantity for analysis; and (b) they represented both good and poor readers (for each treatment, two students were selected from the fourth-grade teachers’ high reading group and two from the low reading group).

The interviews were conducted by the experimenters and two additional graduate students in education who were trained in the interview procedures. The interviewers worked from a prepared script; interview protocols were audio recorded and transcribed later. The interviews consisted of three parts which were ordered such that the most open-ended, unobtrusive tasks were conducted first followed by tasks and questions that were more obviously related to comprehension monitoring.

1. Metacomprehension demonstration. Students read a basal story and conveyed their thoughts, ideas, and reading behaviors at three different times: (a) once before they read the story (from reading the title and a brief introduction, and looking at a picture), (b) a second time halfway through the story, and (c) a third time at the end of the story. The story, titled “The Footbridge,” is an excerpt from the Laura Ingalls Wilder book On the Banks of Plum Creek in which Laura, playing in the hazardous, fast-running waters of Plum Creek, rolls off a footbridge and nearly drowns.

2. General reading questions. The experimenter asked several additional general questions about students’ reading behaviors. Questions such as “Do you ever get mixed up or confused when you read?” and “What do you do to help you understand better when you do get mixed up?” were asked.

3. Forced choice/explanation probes. This task required the students to (a) inspect five sets of statements that were similar to those contained in Posttest 2 (each set consisted of three statements, one of which represented a comprehension monitoring behavior); (b) respond to the question “Which of these things would be the best thing for you to do when you read?” for each set; and (c) justify or explain their choice for each set.

**Procedure**

Pretesting, intervention, and posttesting occurred within a 3-week period. One day was spent on the pretests. The next 10 consecutive days were spent on instruction. Posttesting occurred on the 3 days following instruction. The three experimenters, all of whom were experienced elementary school teachers, conducted all pretesting, instruction, and posttesting (with the exception of Posttest 4, for which two additional graduate students in education assisted with data collection). To
control for teacher influences, a counterbalanced order for instruction and testing was established such that each experimenter worked a total of 4 or 5 days with each group. Instruction in all groups occurred simultaneously during the regularly scheduled language arts period. Instructional groups were assigned randomly to one of three teaching stations, and the experimenters rotated among classes as per the counterbalancing procedures. All instructional lessons were 45 minutes in length, a time determined to be optimal from a pilot study.

Students in each treatment read the same 10 basal reader selections (one story per lesson) in the same sequence. The basal series used in the experiment was different from the school district’s adopted series. All stories were realistic fiction. Grade 3 basal materials were selected in order to minimize decoding difficulties for the lower performing students.

TA instruction. The purpose of TA instruction was to promote students’ comprehension monitoring and self-correction abilities. The TA training was viewed as a vehicle for helping students to acquire control over these abilities; that is, think aloud was a means to an end—improved comprehension monitoring ability—not the end itself. The 10 TA lessons included instruction in the following comprehension monitoring and correction strategies.

Lesson 1: Introduction/self-questioning. The objective of Lesson 1 was to teach students self-questioning as a first step in comprehension monitoring. The concept of self-questioning was introduced by presenting to the students the fictional figure Clark Canine-Super Reporter (CC/SR). CC/SR was modeled after similar metaphors Paris and his colleagues developed for their Informed Strategies for Learning metacognitive instructional program (Paris, 1985; Paris et al., 1984). CC/SR, who appeared throughout the 10 TA group lessons, was presented as a special kind of reporter who interviewed writers. The students were taught to view the role of a reporter (one who interviews people) as being analogous to the role of a reader (one who interviews writers). The students were asked to think of themselves as Pup Reporters, novice writer-interviewers led by CC/SR, whose job it was to interview (i.e., ask questions of) writers whose stories they read. CC/SR was displayed as a large cutout (a canine dressed to look like Clark Kent, i.e., Superman). A chart that compared how reporters and readers are alike was also introduced in Lesson 1. The first pair of items on this chart, which were affixed to the chart during the lesson, indicated that reporters and readers are alike in that they both conduct interviews by asking writers/people questions. Additional ways in which readers and reporters are alike were added to the chart as other comprehension monitoring and correction strategies were introduced in subsequent lessons. For example, in Lesson 2, the two following statements were added to the chart: (a) “Reporters get information from the person they interview and from what they know already” and (b) “Readers get information from the writer and from what
they know already.'" Instruction and application in how to question a writer occurred through the guided reading of the basal story used in Lesson 1.

Lesson 2: Sources of information. Using a modified form of Raphael's (1982) question-answer relationship strategy, students were taught that information can come from ideas in the story and from ideas a student may already possess (i.e., after Pearson & Johnson's, 1978, notion of textually explicit and scriptally implicit information).

Lesson 3: TA introduction: Is the story making sense? Students were introduced to the notion of thinking aloud as they read by being taught to: (a) read short text sections, (b) stop periodically and say out loud what their thoughts were about the story, (c) ask "Is this making sense?", and (d) respond to this question.

Lesson 4: TA review and extension. This was a cumulative review of the first three lessons.

Lesson 5: Predicting, reading, and verifying. Students were taught to use a predict-read-verify strategy (Baumann, 1991); they expressed their predictions and their evaluations of them through think aloud.

Lesson 6: Understanding unstated information. Students were taught to infer unstated information in a story according to a simplified version of the inference categories recommended by Johnson and Johnson (1986); they used think aloud to verbalize what a writer omitted, drawing from story and experience clues.

Lesson 7: Retelling a story. Retelling was presented as a think-aloud strategy for helping a reader to understand and remember a story.

Lesson 8: Rereading and reading on. Students were instructed to stop periodically while reading and ask themselves, "Is this story making sense?" When they responded "no" to this question, they were asked to employ one of two correction strategies to help them clarify meaning, either a rereading strategy or a read-on-and-withhold-judgment strategy.

Lessons 9 and 10: TA/comprehension monitoring application. The final two lessons consisted of review instruction and guided practice of the content of Lessons 1–8.

The Baumann and Schmitt (1986) explicit instruction model was used to teach all TA lessons. This model consists of four steps: (a) What: a description, definition, or example of the comprehension strategy to be taught; (b) Why: an explanation of why the strategy is important and how its acquisition will make students better readers; (c) How: explicit instruction in the use of the strategy, which involves the sequence of verbal explanation, teacher modeling, guided practice, and independent practice; and (d) When: an explanation of the conditions under which the strategy should and should not be used and how to evaluate strategy use. To exemplify how a TA lesson proceeded according to this model, the plan for Lesson 3, "Think-Aloud Introduction: Is the Story Making Sense?", is reproduced in Appendix A.
**DRTA instruction.** DRTA as taught to the children in this experiment consisted of four steps: (a) Students made initial predictions about the story from the title and any pictures on the title page, and predictions were recorded on the chalkboard; (b) students read one-third to one-half of the story and evaluated their initial predictions using the scheme $T = \text{true}$, $F = \text{false}$, $PT = \text{partly true}$, $PF = \text{partly false}$, and $NM = \text{not mentioned}$; (c) students made additional predictions; and (d) steps b and c were repeated once or twice more, depending upon the length of the story.

**DRA instruction.** Students read the same stories according to the following steps which were taken from the basal reader teacher edition: (a) introducing the new vocabulary; (b) activating and/or providing background knowledge; (c) guiding the students’ reading of the selection through purpose-setting and comprehension questions; and (d) discussing the selection by returning to the purpose, summarizing the selection, and responding to comprehension questions.

**Fidelity of Treatments.** To establish fidelity of treatments, two undergraduate education majors each observed the researchers teach two lessons and listened to audiotapes of two additional lessons (i.e., a total of eight lesson evaluations). These students were independent in their evaluations to the extent that they were blind to the purpose of the experiment and had no indication as to which group was experimental, comparative treatment, and control. The evaluators were directed to assess fidelity of treatments according to three criteria, which were presented as questions: (a) Were the experimenters faithful to the content and organization of the lessons by adhering to the lesson plans? (b) Was there interexperimenter consistency, that is, was each experimenter’s instruction compatible with that of the others? and (c) Were the experimenters unbiased in their instruction, that is, did they display equal amounts of enthusiasm, administer equal amounts of praise, and so forth, with each group?

For example, for criterion “a,” each evaluator compared the audio records of two lessons to the corresponding lesson plans to determine if the experimenters followed the planned content and in the proper sequence. For instance, for TA Lesson 3, “Think-Aloud Introduction” (see Appendix A), one evaluator checked to make certain that the experimenter: (a) communicated the What and Why information, (b) provided the think-aloud rules and modeled them, (c) engaged the students in guided and independent practice, (d) and infused the When information into the guided and independent practice.

The student evaluators analyzed the lessons independently using the three criteria and recorded their evaluative comments on paper. They then met and compared their assessments of the experimenters’ fidelity relative to the criteria and finally reported their analyses to the experimenters. The evaluators’ analyses documented that fidelity of treatment was consistently high for all three evaluation criteria.

**Scoring and Data Analyses: Posttests 1–3.** Posttests 1–3 were scored by com-
paring students' responses to prepared answer keys. Total possible number of correct items for Posttests 1–3 were 16, 18, and 56, respectively.

For Posttests 1 and 3, which evaluated aspects of comprehension monitoring, a multiple analysis of covariance (MANCOVA) was conducted, using Pretests 1 and 2 as covariates. Two orthogonal contrasts (within a univariate analysis of covariance model) were planned as possible follow-up procedures to a significant omnibus MANCOVA. The contrasts were designed to address the two basic research questions. Specifically, an Effect of Instruction contrast compared training in TA and DRTA to the DRA group (i.e., TA + DRTA vs. 2 × DRA), and an Intensity of Instruction contrast compared instruction in TA to instruction in DRTA (i.e., TA vs. DRTA).

Because Posttest 2 was different substantively from Posttests 1 and 3 (i.e., it elicited self report data on students' awareness of and sensitivity to comprehension monitoring, whereas Posttests 1 and 3 evaluated their actual comprehension monitoring performance), it was not included in the MANCOVA. However, the same planned contrasts were conducted for Posttest 2, again using Pretests 1 and 2 as covariates.

Scoring and Data Analyses: Posttest 4. The 12 transcribed interviews were analyzed by two of the researchers for the presence of statements that indicated students' awareness of or use of comprehension monitoring and think-aloud behaviors. The researchers were blind to students' group membership during protocol analysis (adapted from Patton, 1980), which proceeded as follows:

1. Protocols were read independently several times to establish a set of categories of responses (e.g., ask questions, predict, retell). The researchers met, compared categories of responses, and agreed upon a final set. Next, the researchers independently categorized and tallied student responses to Part 1 data (metacomprehension demonstration) in two ways: according to those behaviors the students said they did as they read and those behaviors the students actually demonstrated as they read. For example, in response to the Part 1 question, "What were the things you were doing as you read," a student might respond, "I was making guesses." This response would be classified as a student saying she was making a prediction. However, if another student responded to the same question, "I guessed that Laura was going to drown after she rolled off the plank into the creek," this response would be classified as a student actually demonstrating the behavior of prediction.

2. The researchers independently analyzed Part 2 data (general reading questions) and categorized student responses. All these responses were classified as indications of what students said they did as they read since questions queried them about general reading comprehension strategies.

3. Part 3 data (forced choice/explanation probes) were scored objectively, and student justifications were categorized.
4. Interscorer reliability for categorizing student responses was calculated (91% agreement score across all measures).

5. The researchers collectively analyzed the composite profiles and predicted group membership for each of the 12 students.

RESULTS

Posttests 1–3

Table 1 presents mean observed scores, standard deviations for mean observed scores, and mean adjusted scores for the Posttests 1, 2, and 3 by treatment group.

The MANCOVA for Posttests 1 and 3 was significant according to Wilks’ lambda, $F(4, 120) = 5.75$, $p<.001$. For Posttest 1, the error detection test, both planned contrasts were significant. The Effect of Instruction contrast demonstrated that TA and DRTA students outperformed DRA students, $F(1, 61) = 20.14$, $p<.001$; the Intensity of Instruction contrast indicated that DRTA students outperformed TA students, $F(1, 61) = 4.90$, $p<.030$.

For Posttest 3, the Degrees of Reading Power, the Effect of Instruction contrast indicated that the combined effect of TA and DRTA instruction was superior to DRA instruction, $F(1, 61) = 7.00$, $p<.010$; the Intensity of Instruction contrast failed to attain statistical significance, $F(1, 61) = 1.65$, $p<.203$.

For Posttest 2, the comprehension monitoring questionnaire, both planned contrasts attained statistical significance. As for Posttests 1 and 3, the Effect of Instruction contrast indicated that the combined effect of TA and DRTA instruction was superior to DRA instruction, $F(1, 61) = 9.68$, $p<.003$; however, the Intensity of Instruction contrast indicated that TA students outperformed DRTA students, $F(1, 61) = 10.07$, $p<.002$.

Posttest 4

Metacognition. Table 2 presents frequencies of behaviors the 12 interviewed students said they used (Said) and those they actually demonstrated (Did) as they read “The Footbridge” and responded to the experimenters’ probes. A comparison of total responses indicates that the TA students reported using more comprehension monitoring behaviors than students in the other groups (Said: TA = 32, DRTA = 14, DRA = 5), and they actually demonstrated greater use of them (Did: TA = 41, DRTA = 34, DRA = 24). Thus, overall, TA students engaged in more comprehension monitoring behaviors (73 total instances, $M = 18.25$) than either DRTA students (48 total behaviors, $M = 12.00$) or DRA students (29 total behaviors, $M = 7.25$).

Prediction was the most frequently occurring type of behavior, and DRTA students did the most predicting (Total Said/Did Prediction: TA = 27, DRTA = 35,
Table 1

*Mean Total Correct Scores by Treatment Group for Posttests 1–3*

<table>
<thead>
<tr>
<th></th>
<th>TA</th>
<th>DRTA</th>
<th>DRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest 1: Error detection test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M^a$</td>
<td>7.773</td>
<td>9.773</td>
<td>6.682</td>
</tr>
<tr>
<td>$SD^b$</td>
<td>3.927</td>
<td>2.724</td>
<td>2.767</td>
</tr>
<tr>
<td>$M^c$</td>
<td>8.222</td>
<td>9.185</td>
<td>6.190</td>
</tr>
<tr>
<td>Posttest 2: Comprehension monitoring questionnaire</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M^a$</td>
<td>8.364</td>
<td>6.227</td>
<td>5.545</td>
</tr>
<tr>
<td>$SD^b$</td>
<td>2.904</td>
<td>2.092</td>
<td>2.071</td>
</tr>
<tr>
<td>$M^c$</td>
<td>8.461</td>
<td>6.237</td>
<td>5.439</td>
</tr>
<tr>
<td>Posttest 3: Degrees of Reading Power</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M^a$</td>
<td>43.455</td>
<td>46.636</td>
<td>42.046</td>
</tr>
<tr>
<td>$SD^b$</td>
<td>7.860</td>
<td>7.644</td>
<td>6.615</td>
</tr>
<tr>
<td>$M^c$</td>
<td>44.289</td>
<td>46.715</td>
<td>41.133</td>
</tr>
</tbody>
</table>

*Note.* Maximum possible scores: Posttest 1 = 16, Posttest 2 = 18, Posttest 3 = 56.

$^a$Mean observed scores. $^b$Standard deviations for mean observed scores. $^c$Mean scores adjusted for total correct on Pretests 1 and 2.

DRA = 11). However, TA students not only reported or demonstrated more total behaviors, they displayed a greater range of behaviors. To document this, when the prediction values are excluded from Table 2, TA students still reported or demonstrated 46 other comprehension monitoring behaviors ($M = 11.5$), whereas the values were significantly lower for DRTA students (13 total other behaviors, $M = 3.25$) and DRA students (18 total other behaviors, $M = 4.5$). In short, the metacomprehension demonstration indicated that students learned their lessons well: TA students reported or demonstrated a variety of metacomprehension behaviors; DRTA students made many predictions (73% of all said/did behaviors were predictions); and DRA students did little to demonstrate either an awareness or ability to engage in comprehension monitoring. Protocol excerpts from the metacomprehension demonstration presented in Table 3 support these conclusions.

**General Reading Questions.** Table 4 displays the types of responses (and their frequencies) students made to the general reading questions (e.g., “What do you do to help you understand when you read?”). TA students made more total responses (25) compared to DRTA students (11) and DRA students (15). Further, TA students reported more comprehension monitoring behaviors. For example, for the first five behaviors listed on Table 4, all of which are aspects of comprehension...
Table 2

Said and Did Responses from Posttest 4, Metacomprehension Demonstration

<table>
<thead>
<tr>
<th>Category</th>
<th>AQ</th>
<th>AnQ</th>
<th>Pre</th>
<th>Ver</th>
<th>MS</th>
<th>Sum</th>
<th>Par</th>
<th>Rer</th>
<th>PK</th>
<th>Ret</th>
<th>Inf</th>
<th>Tot</th>
<th>TotS/D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Said</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>32</td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>Did</td>
<td>8</td>
<td>21</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Said</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>48</td>
</tr>
<tr>
<td>Did</td>
<td>1</td>
<td>26</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Said</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Did</td>
<td>4</td>
<td>11</td>
<td>1</td>
<td>6</td>
<td></td>
<td>1</td>
<td>1</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: AQ = ask questions; AnQ = answer questions; Pre = predict; Ver = verify; MS = asks “Is this making sense?”; Sum = summarize; Par = paraphrase; Rer = reread; PK = prior knowledge; Ret = retell, Inf = inference; Tot = row total; TotS/D = combined said/did total.
Table 3

Responses to Experimenter Questions and Follow-Up Probes* from Posttest 4, Metacomprehension Demonstration

<table>
<thead>
<tr>
<th>Think-Aloud Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tom</strong>:</td>
<td>You can reread it [story] sometimes and then think about it, like trying to recite what happened in the story.</td>
</tr>
<tr>
<td><strong>Sam</strong>:</td>
<td>Retell it . . . Could ask if you think it made sense.</td>
</tr>
<tr>
<td><strong>Kim</strong>:</td>
<td>When I read I think, is this making sense? I might . . . ask questions about the story and reread or retell the story . . . I was asking myself is this making sense and I was asking if like do I think what would happen next without reading the next page.</td>
</tr>
<tr>
<td><strong>Sam</strong>:</td>
<td>[I] think about the title, “The Footbridge.” [I] think since it’s by Laura Ingalls Wilder—she always writes about times back in the pioneers setting—that maybe the footbridge is a—maybe she made a, like a bridge across a little creek with rocks or something.</td>
</tr>
<tr>
<td><strong>Ann</strong>:</td>
<td>I was asking questions, and I asked questions like “Why did she go to the creek when her mother told her not to?” And “Why did Laura take her shoes and socks off when she knew the creek was going to be rocky and muddy on the bottom?”</td>
</tr>
<tr>
<td><strong>Kim</strong>:</td>
<td>I was asking myself is this making sense, and I was asking if like do I think what would happen next without reading the next page—just reading that [the present] page. [Exp.: Can you tell me a bit more about this?] She’ll probably go down there again and play when the water’s down and when it’s not so high and when it’s not so like roaring and stuff.</td>
</tr>
<tr>
<td><strong>Sam</strong>:</td>
<td>I retold what I read the first time to [page] 193 as I was reading the last part of the story to see if it would make sense.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DRTA Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Al</strong>:</td>
<td>After I started reading that last part, [I thought] that she was going to get out after all. After I started reading that page, I could see that she got back on the plank. And then I was thinking that she would probably get in trouble when she came home.</td>
</tr>
<tr>
<td><strong>Sue</strong>:</td>
<td>I was thinking about something would happen to her and at the creek. Her mom would come out looking for her or if she turned out missing or something.</td>
</tr>
<tr>
<td><strong>Jim</strong>:</td>
<td>I was thinking about if she could grab the plank and force it down so it would block her from the waves and she wouldn’t be in trouble and then she could just climb right out of the thing—sorta crawl up . . . . I was sort of thinking about the future, but I was just remembering the past and thinking about the past and present and mixed those together and I came up with what would happen in the future.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DRA Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lynn</strong>:</td>
<td>Oh, trying to stop at every period and trying to pause at the commas . . . . I was trying to read loud like instead of talking real soft and you couldn’t hear me. [Exp.: What else? Anything else you did or thought about as you were reading this section?] Not really.</td>
</tr>
</tbody>
</table>
Table 3 (Continued)

Exp.: What were you doing or thinking about as you read this part of the story?
Kate: Nothing.
Exp.: Nothing? What kind of ideas did you have as you read?
Kate: That her mom was very nice and understood that it could have killed her.
Exp.: Any other ideas you had?
Kate: She was nice.
Exp.: Anything else?
Kate: No.

*Experiment questions and probes included statements such as the following: "What do you do before you get ready to read?" "What were you thinking about as you read?" "What were the things you were doing as you read?" "Can you tell me any more about this?" "Is there anything else you did or thought about?" *Students*’ names are fictitious.

Table 4

Types and Frequencies of Responses to Posttest 4, Follow-Up General Reading Questions

<table>
<thead>
<tr>
<th>Behavior</th>
<th>TA</th>
<th>DRTA</th>
<th>DRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read on</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Reread</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ask, &quot;making sense?&quot;</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ask questions</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retell/recall/paraphrase</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Use context</td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Use pictures</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Remember everything</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read every line</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Use SQ3R</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use dictionary/glossary</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ask friend/parent</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Let someone else read</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read it aloud</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read someplace quiet</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Read whole story again</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use phonics</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Read carefully</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Act like I’m there</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Total responses</td>
<td>25</td>
<td>11</td>
<td>15</td>
</tr>
</tbody>
</table>
monitoring, TA students noted them 11 times compared to 6 for DRTA and 2 for DRA students. Finally, only DRA students mentioned reading behaviors that were not focused on acquiring meaning in some fashion (e.g., "Use phonics").

The following sample responses to the general reading questions corroborate these quantitative data:

*Ann* (TA): I ask all the time is this making sense, and like this week, I checked out a book and I looked at the title and I didn’t really understand it. But once I got reading it, it made sense.

*Jim* (DRTA): I can think about it [a confusing part of a story] . . . and think about what happened in the past with that and then the present and then also the future. And I can sort of piece together in the puzzle and I can find the missing part so I can understand what they’re doing.

*Kate* (DRA): Try to pronounce words that I can’t. And words that I don’t know what they mean look up in the dictionary.

**Forced Choice/Explanation Probes.** Table 5 presents student response data by group for the five sets of three-option, forced-choice items. For example, three of four TA students selected the “ask myself, is it making sense?” option for Set 1. As can be seen, TA students more frequently selected the comprehension monitoring option (13 out of a possible total of 20 items) than DRTA students (10/20) and DRA students (5/20). Furthermore, a tabulation of the noncomprehension monitoring responses indicated that DRA students selected the phonics or oral reading option 12 times, whereas TA and DRTA students selected these options only 6 and 7 times, respectively.

When probed to justify their selections, students’ responses were generally consistent with the option they selected:

**Table 5**

| Comprehension Monitoring Behaviors Selected for Posttest 4, Follow-Up Forced Choice Questions |
|---|---|---|---|---|---|---|
| Group | Set 1 Making sense? | Set 2 Make guesses | Set 3 Retell | Set 4 Ask questions | Set 5 Think aloud | ΣX  |
| Think Aloud | 3 | 3 | 4 | 2 | 1 | 13 | 3.25 |
| DRTA | 2 | 2 | 3 | 2 | 1 | 10 | 2.5 |
| DRA | 2 | 1 | 2 | 0 | 0 | 5 | 1.25 |

*Note.* Each set contained three options, one of which involved a comprehension monitoring behavior. These data report numbers of students per group (out of a total of four subjects per group) who selected the comprehension monitoring option.
Tom (TA, selected “Making sense?”): Because if I ask if things are making sense, it helps me remember what has happened and I don’t forget. Because sometimes I forget what has happened and I get mixed up.

Jim (DRTA, selected “Make guesses”): Helps understand the story better . . . . It’s sort of fun to figure out what happens in the future.

Mark (DRA, selected “Use phonics”): It might make the story make sense more if you used phonics. If you don’t get a word, you can sound it out and use the consonant blends and long and short vowels.

The final step in interview protocol analyses involved predicting group membership for the 12 students interviewed. The two researchers who analyzed the protocols made these predictions independently, and they were able to successfully predict group membership for 11 of 12 (92%) students. Both researchers misidentified Mark, a DRA student who demonstrated a number of prediction behaviors, as being a member of the DRTA group.

DISCUSSION

This experiment was conducted to address two questions: (a) Does TA training or the use of the DRTA enhance students’ ability to monitor their reading comprehension when compared to a traditional DRA? and (b) Is there any difference between TA instruction and DRTA instruction in promoting students’ comprehension monitoring ability? This discussion is organized by addressing each question in turn.

Regarding the first question, the data clearly indicate that either TA training or use of the DRTA is a superior method for promoting students’ comprehension monitoring abilities. The significant Effect of Instruction contrasts for Posttests 1–3, which favor the combined effects of TA and DRTA over DRA, support this conclusion. This trend is also found in the quantitative and qualitative analyses of student interview data, as TA and DRTA students were more likely than DRA students to report or demonstrate using various comprehension monitoring behaviors. In other words, it appears that some kind of intervention, be it TA or DRTA, is superior to a traditional DRA for promoting students’ comprehension monitoring awareness and skill.

This finding was not unexpected, of course, as there is ample extant research supporting the efficacy of cognitive strategy training during reading as a means to enhance students’ comprehension monitoring abilities (e.g., Paris et al., 1988). Further, the consistently poor performance of the DRA group reinforces the notion that didactic, noninteractive instruction fails to promote students’ comprehension monitoring abilities. With the exception of students who may intuit or acquire comprehension monitoring abilities independent of formal instruction (e.g., Mark, a DRA student in this experiment), it is clear that teacher-led instruction in strategies like TA and DRTA, or others which cultivate student responsibility and control
for reading comprehension (e.g., Miller, 1985, 1987; Miller et al., 1987; Schmitt, 1988), are effective for developing comprehension monitoring skills.

Results that address the second question, the relative effectiveness of TA versus DRTA, are more ambiguous. The Intensity of Instruction contrast for the comprehension monitoring questionnaire (Posttest 2) indicates that TA students had greater awareness of comprehension monitoring and correction strategies than DRTA students. Likewise, interview responses (Posttest 4) indicate that TA students both reported and demonstrated using a greater depth and breadth of comprehension monitoring behaviors than their DRTA counterparts. The experimenters' ability to successfully predict group membership on the basis of interview protocols provides further support for the assertion that TA training was more effective than DRTA in promoting comprehension monitoring.

However, the whole-sample comprehension performance measures (Posttests 1 and 3) suggest that DRTA may be a more powerful strategy than TA for promoting comprehension monitoring. DRTA students outperformed TA students on the error detection test (Posttest 1), and the same trend, though not statistically reliable, was evident for the Degrees of Reading Power cloze test (Posttest 3).

How might these equivocal findings be explained? One simple but plausible explanation is that prediction plays a more central role in comprehension monitoring than previously recognized. Although comprehension monitoring training studies have routinely included prediction as one element of instruction (e.g., Bereiter & Bird, 1985; Palincsar & Brown, 1984; Paris et al., 1984; Schmitt, 1988), none, prior to the present study, has isolated its singular effect on comprehension monitoring. Thus, it may be that prediction is one of the more salient components of comprehension monitoring and that intensive instruction and practice in prediction alone result in significantly enhanced comprehension monitoring abilities. Future experiments that replicate and extend the present one are required to determine the relative power of DRTA as a technique for fostering comprehension monitoring.

A second possible explanation involves strategy complexity and the duration of instruction. In this experiment, seven comprehension monitoring strategies were presented to the students in 10 instructional sessions. In contrast, for example, Paris et al. (1984) distributed instruction in 13 strategies over 4 months. Thus, the relatively high number of comprehension monitoring strategies that were included in the short-term TA instruction may have diminished the effectiveness of the treatment. TA students may have been overwhelmed or confused by the presentation of multiple strategies in such a short period of time. On the other hand, significant treatment effects have been reported for experiments having relatively short-term intervention periods (e.g., Bereiter & Bird, 1985; Miller, 1985, 1987; Palincsar & Brown, 1984; Schmitt, 1988), and several of them involved instruction in complex sets of strategies. Therefore, additional research is needed to determine how strategy complexity, the duration of training, and perhaps even the pacing of instruction affect the acquisition of comprehension monitoring skills.
A third possible explanation for the equivocal data involves issues related to the measurement of metacognition and comprehension monitoring (see Garner, 1987, for a lengthy discussion). Because it is difficult to directly measure covert comprehension monitoring behaviors such as self-questioning, paraphrasing, predicting, and the like, their presence must be inferred from overt behaviors. Although there are theoretical and empirical bases for the dependent measures used in this experiment, the windows they provide for observing cognitive processes are not free from distortions. For example, Winograd and Johnston (1982) have identified various limitations of the error detection paradigm for assessing comprehension monitoring. Other researchers note that differing task demands and purposes for reading, as well as the type of text inconsistency itself, all influence readers’ responses to textual error or inconsistencies (Baker 1985; Moore & Zabrucky, 1989; Zabrucky & Moore, 1989). In addition, Baker and Brown (1984a) point out that the disruptions of a cloze test make it somewhat removed from natural reading tasks. Likewise, self-report and verbal report measures, such as the comprehension monitoring questionnaire and the student interviews, are far from perfect indices of comprehension monitoring behavior (cf. Afflerbach & Johnston, 1984; Baker & Brown, 1984b; Garner, 1987). For example, a child’s verbal facility strongly influences the quantity and quality of verbal self-reports (Cavanaugh & Perlmutter, 1982). Additional research and experimentation are required to sort out what comprehension monitoring behaviors are and are not assessed by various measures.

One means to minimize the measurement limitations of extant methods for assessing comprehension monitoring is to include multiple measures in an experiment, measures that, one hopes, converge toward answers or conclusions (Kail & Bisanz, 1982). Although the windows on metacognitive processing provided by the measures selected for this experiment do not provide a perfectly focused view of comprehension monitoring, we believe they do converge sufficiently to warrant several conclusions.

**Conclusions**

Three conclusions are drawn from the results of the present study. First, the experiment demonstrates that explicit, teacher-led instruction in think aloud is an effective means to enhance students’ comprehension monitoring abilities. This finding supports Bereiter and Bird’s (1985) reported success in using think alouds, and it also provides empirical support for the efficacy of the Baumann and Schmitt (1986) explicit instruction model. The significant amount of verbal explanation and modeling by the teacher provided by this model also reinforces Bereiter and Bird’s contention that a combination of these behaviors is required for students to acquire facility with think-aloud tasks.

Second, the experiment demonstrates that DRTA is also an effective means to promote students’ comprehension monitoring abilities. Although DRTA students
did not report or demonstrate the variety of comprehension monitoring behaviors their TA counterparts did, their performance on measures believed to tap dimensions of comprehension monitoring was equivalent to or exceeded that of TA-trained subjects. It is also interesting to note that DRTA-instructed students tended to generalize their experience in learning to predict to other non instructed comprehension monitoring behaviors (i.e., they reported and demonstrated using comprehension monitoring behaviors such as visualizing, rereading, and paraphrasing).

Third, we conclude from the performance of our instructed-control subjects that teachers must engage students interactively with selections in some fashion in order to promote comprehension monitoring behaviors. DRA experiences seemed to reinforce for most students that reading is basically a literal, if not word-level, task.

No conclusions about the relative effectiveness of TA versus DRTA can be drawn from this experiment. Additional intervention research is required to explain the somewhat surprising effectiveness of DRTA. As we have postulated, it may be that prediction is such an integral component to comprehension monitoring that training in it alone is sufficient to enhance comprehension monitoring. If the positive effects of DRTA training on students' comprehension monitoring behaviors are replicated in subsequent studies, then the efficiency of TA instruction might be questioned, for it is simpler for teachers and students to adopt the DRTA than the more elaborate sequence of TA instruction.

Limitations

These conclusions need to be tempered, however, by several limitations of the experiment. First, the instruction was conducted by the experimenters themselves. Doing so enhanced internal validity of the experiment but diminished its external validity. It remains to be determined if the results of this experiment can be replicated when regular classroom teachers implement the interventions.

The short duration of the experiment may be another limitation. If instruction is extended over a longer period of time, students may be able to assimilate comprehension monitoring behaviors into their repertoire of reading strategies more readily (e.g., Paris et al., 1984). Extending comprehension monitoring training into other reading tasks (e.g., content reading assignments) would also help students internalize these complex behaviors. On the other hand, since most strategy instruction in schools is presented in units of instruction or from basal materials which cluster instruction within short periods of time, there is some ecological validity to the relatively brief, intensive nature of the instruction provided in this experiment, even though pedagogically instruction of longer duration may be more sensible.

The pretests and posttests used in the experiment also present potential limitations. Although multiple dependent measures were selected, all with an empirical or theoretical rationale for their selection, they only tap certain dimensions of
comprehension monitoring awareness and performance. Thus, the conclusions 
drawn from this experiment are limited to the manner in which comprehension 
monitoring is defined operationally by those measures.

Additionally, the reliability of the assessments used in the experiment are 
modest. The reliabilities calculated according to the Kuder-Richardson 21 formula 
for Pretest 1 (.62), Posttest 1 (.70), and Posttest 3 (.85) are acceptable when 
aggregate scores are examined, as in the present study (Salvia & Ysseldyke, 1988, 
indicate that reliability coefficients of .60 or greater are acceptable for group deci-
sion making). However, the reliabilities for Pretest 2 (.36) and Posttest 2 (.42), 
the comprehension monitoring questionnaire, are low, a function of the rather low 
viances for students’ performance on these instruments. These coefficients were 
somewhat surprising since other instruments after which these assessments were 
modeled were shown to be more reliable (e.g., Lonberger, 1988; Paris et al., 1984; 
Paris & Jacobs, 1984; Schmitt, 1988). In their review of experimental studies on 
comprehension strategy instruction, Lysynchuk et al. (1989) found that experiment-
ers reported reliabilities in only 13 of the 37 studies (35%) they examined, so it is 
difficult to assess how the reliability indices calculated in this experiment compare 
to that of instruments developed for use in most other comprehension strategy 
instructional studies. Nevertheless, conclusions drawn from the comprehension 
monitoring questionnaire in the present study should be interpreted cautiously.

Finally, the results of this experiment should not be generalized beyond the 
scope of the subjects, methods, and tasks that were employed. Future research on 
the TA and DRTA strategies with other populations in other contexts is needed to 
determine if these results are generalizable.

In conclusion, one finds increasing emphasis placed on strategic reading behav-
iors, including comprehension monitoring, in professional materials for teachers 
(e.g., Irwin & Baker, 1989; Wilson & Gambrell, 1988), and in instructional materi-
als for children (e.g., Alvermann, Bridge, Schmidt, Searfoss, & Winograd, 1991; 
Pearson et al., 1991). However, we do not assume that teachers routinely teach 
children metacognitive behaviors such as comprehension monitoring. In fact, 
Schmitt and Baumann (1990) reported that teachers tended to do little spontane-
ously to promote children’s metacognitive awareness and ability during conven-
tional reading instruction lessons. The present study provides additional evidence 
that teachers now have at their disposal a variety of techniques to instruct students 
in various metacognitive and comprehension monitoring abilities.

REFERENCES


Think-Aloud Instruction


Moore, Miller, Miller, Lysynchuk, Miller, Kail, Irwin, Hayes, Garner, DRP, Hare, Ericsson, 16, 8, 16, 68,
S, comprehension-monitoring, comprehensio,
terly, average, &
G, G, R, average,
R, average,
R
24,
G, 24, &
L, 'e
effect
Jr., g, g,
(1985).
Kirby e 458-470 M.
f Singer, Smith, (1986)
H, 12, e 12,
effect
p, n,
i nfluence, 21,
D, t theron, e 21,
E
I
H,
J
A.
(1986)

Hare, V. C., & Smith, D. E. (1982). Reading to remember: Studies of metacognitive reading skills in
protocol analysis. In P. Mosenthal, L. Tamor, & S. A. Walsmley (Eds.), Research on writing:
Principles and methods (pp. 207–220). New York: Longman.
processing abilities of 4th and 6th graders. Paper presented at the annual meeting of the Eastern
Education Association, Savannah, GA.
analysis of experimental studies of comprehension strategy instruction. Reading Research Quar-
terly, 24, 458-470.
reading-comprehension performance for poor comprehenders. The Elementary School Journal,
90, 469–484.
Press.
Miller, G. E. (1985). The effects of general and specific self-instruction training on children’s compre-
of Reading Behavior, 21, 295–307.
Nist, S. L., & Kirby, K. (1986). Teaching comprehension and study strategies through modeling and
thinking aloud. Reading Research and Instruction, 25, 256–264.
Research Quarterly, 12, 654–675.
comprehension-monitoring activities. Cognition and Instruction, 2, 117–175.
Think-Aloud Instruction


Plan for Lesson 3: "Think-Aloud Introduction: Is the Story Making Sense?"

**Objective:** To introduce to students the notion of think aloud as a vehicle to monitor and aid comprehension; to teach students to monitor comprehension while reading by asking themselves, "Is this making sense?"

**WHAT**

Begin by briefly reviewing what was taught in Lessons 1 and 2, using the "How Readers and Reporters are Alike" instructional chart. Then inform students that today they will learn to improve their understanding of a story by saying out loud what goes on in their minds as they read. They will do this by stopping periodically and asking themselves, "Is this making sense?"

**WHY**

Explain that the think aloud and self-questioning strategies are powerful ways to determine if something is making sense and to help a reader understand better when something is confusing.

APPENDIX A

Plan for Lesson 3: "Think-Aloud Introduction: Is the Story Making Sense?"

**Objective:** To introduce to students the notion of think aloud as a vehicle to monitor and aid comprehension; to teach students to monitor comprehension while reading by asking themselves, "Is this making sense?"

**WHAT**

Begin by briefly reviewing what was taught in Lessons 1 and 2, using the "How Readers and Reporters are Alike" instructional chart. Then inform students that today they will learn to improve their understanding of a story by saying out loud what goes on in their minds as they read. They will do this by stopping periodically and asking themselves, "Is this making sense?"

**WHY**

Explain that the think aloud and self-questioning strategies are powerful ways to determine if something is making sense and to help a reader understand better when something is confusing.
Think-Aloud Instruction

HOW

Explanation
* Explain that thinking aloud is saying what is going on in your mind as you try to understand a story or solve a problem.
* Ask the students if they ever do this, for example, when doing a hard math problem or reading difficult directions.
* Put the following think-aloud “rules” on the board:

To Help Me Read and Understand:
1. Say out loud what it going on in my mind as I read.
2. Ask myself, “Is this making sense?”

* To demonstrate what thinking aloud is like, write the following verbal and mathematical analogies on the board and think aloud as you solve them. During the analogy solving process, ask yourself, “Is this making sense?”

<table>
<thead>
<tr>
<th>dog: bark</th>
<th>cat: ?</th>
<th>oink, meow, puppy, feline</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:4</td>
<td>5:</td>
<td>5, 10, 2, 8</td>
</tr>
</tbody>
</table>

* Have volunteers then think aloud while solving the following analogies.

<table>
<thead>
<tr>
<th>puddle: lake</th>
<th>hill: ?</th>
<th>valley, ocean, mountain, bump</th>
</tr>
</thead>
<tbody>
<tr>
<td>20:10 50:</td>
<td>25, 150, 30, 100</td>
<td></td>
</tr>
</tbody>
</table>

* Explain that thinking aloud can also be done during reading.
* Affix the Lesson 3 sentences to the “How Readers and Reporters are Alike” chart.
* Inform the students that it is important while reading to think aloud and to ask oneself, “Is this making sense?”
* Explain that saying “Is this making sense?” is an important part of thinking aloud because it forces a reader to check to see if a story is being understood.

Modeling
* Explain that you will show them how to think aloud by doing so for the first part of today’s story, “Food’s on the Table.”

* Begin with the title. Talk through what you think this story might be about from the title; end with a question or prediction drawn from the title (include the two sources of knowledge taught about yesterday).
* Proceed with the first sentence, and continue to model think aloud. Gradually increase the length of text that you use for demonstration. Be sure to return to questions and predictions that you made previously.
* Include within each think aloud segment the question “Is this making sense?” and answer it honestly.

Guided Practice
* Begin to include the students in the modeling examples by asking them: “What ideas are going through your mind?” Elicit several responses to demonstrate that there are no “right” think alouds and that it is important for them to ask themselves questions as they read. Review the CC/SR comparison for this lesson.
* Ask the students to try this out with short text sections. Have a volunteer read the next sentence/paragraph aloud, stop after reading it, think aloud, and ask, “Is this making sense?”
* Have several volunteers do this, gradually increasing the length of text sections.
Independent Practice
*Have the students try the think aloud on their own. Have them read short sections and think aloud as they go. Have them share and discuss their responses.
*Next have pairs of students work together alternately reading paragraphs and thinking aloud for one another.

WHEN
During the guided or independent practice, provide the students the following conditional knowledge:
*Thinking aloud will help you to understand stories and informational writing.
*Use think aloud when you get confused when you read; it also might help you study for a test, for example, when studying from your science or social studies book.
*Asking “Is this making sense?” will help you to decide if you are understanding what you are reading; if you are not, you can try thinking out loud to help you understand better.