DOCUMENT RESUME

ED 285 123 CS 008 834

AUTHOR Palincsar, Annemarie Sullivan
TITLE Collaborating for Collaborative Learning of Text Comprehension.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Apr 87
GRANT 400-81-0030
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987). Contains small print.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Experimental Teaching; Inservice Teacher Education; Middle Schools; *Reading Comprehension; Reading Improvement; *Reading Instruction; Reading Research; Reading Strategies; *Remedial Reading; Remedial Teachers; Teaching Methods
IDENTIFIERS *Collaborative Learning; *Reciprocal Teaching

ABSTRACT

To determine the effect of training in reciprocal teaching on collaborative learning, a study of six middle school remedial reading classes was undertaken. Reciprocal teaching is a method in which the teacher participates both as a leader and as respondent. Before the experiment began, the teachers were trained for a day in reciprocal teaching and introduced to principles of collaborative learning, such as the joint role of student and teacher in acquiring comprehension strategies, the role of teacher modeling, and the importance of allowing students to lead discussions. Participants were 63 experimental and 66 control students. After pretesting, the teachers in the experimental classes introduced collaborative learning, and four reading comprehension strategies. Twenty days of reciprocal teaching followed, with each teacher implementing new stages in the process at the same time. Control groups used traditional basal reading instruction. During a 5-day maintenance phase, subjects completed daily assessment measures, after which students in both groups completed summarization and questioning transfer measures as a posttest. Three of the teachers were more comfortable with reciprocal teaching than the others, although assistance from leaders helped all six groups receive the same quality of instruction. Results revealed a significant improvement in experimental students' accuracy with recall and interpretation of material they read independently. Experimental subjects scored significantly higher on the posttests as well. It is also important that the students enjoyed collaborative learning, made gains in metacognition, and exhibited few behavior problems in the unstructured setting. (References and four tables of data are appended.) (SKC)

Collaborating for Collaborative Learning of Text Comprehension

Annemarie Sullivan Palincsar
Michigan State University

Paper presented at the annual conference of the American Educational Research Association, Washington, D.C., April, 1987

Acknowledgements

The research reported in this paper was supported by a grant, NIE 400-81-0030, from the National Institute of Education. The author wishes to thank the teacher-collaborators from the Springfield, Illinois schools who participated in this study as well as their coordinator, Ms. Kathryn Ransom. Many thanks to Melinda Samuel for her assistance with data collection, Deborah Dunn for scoring the data, Marilyn Ryan for transcribing a number of the audio-tapes, and Yvonne David for preparing the figures and tables. Lynne Webber was the statistical consultant and I am very grateful for that

Introduction

A group of seventh graders participating in the study I will present today had the following discussion:

Cassandra: Now I'll summarize. This paragraph was about why scientists go to study pollution, and about air and water pollution in some parts of the world, and the winds and ocean currents, and scientists are spread all over the world .

(Big sigh)

Okay. This paragraph was about scientists come to study pollution and air and water pollution from parts of the world, and winds and ocean currents, and. scientists.

Billy: The sentence was, "air and water pollution from all parts of the world comes here in wind and water currents " That makes sense? (Disbelievingly)

Students: NO!

Anthony: That sentence - I think what they did was form a sentence - and ..I think what they're talking about is that scientists come here

Cassandra: To Antarctica

Anthony: from other parts of the world to study the pollution. I'm not sure where the air and water pollution and all this comes from - I'm not sure how the sentence goes together, but I don't think it's correct there

Billy: That's where the currents come in and the wind they bring the pollution.

The preceding discussion was taken from a reciprocal teaching lesson Reciprocal teaching is a form of cooperative learning Slavin (1983) has suggested that, at the heart of cooperative learning, is collaboration among learners. This paper focuses on collaboration in the instruction and learning of text comprehension There are three levels of collaboration

represented in this study: collaboration among the teachers and researchers collaboration among the teachers and their students, and collaboration among the students. There is a fourth level of collaboration that I want to be sure to mention and that is the collaboration between Ann Brown and I in the completion of this research

I will begin with a description of reciprocal teaching, which is a discussion conducted by the members of an instructional group with the teacher participating as both a leader and respondent The dialogue is structured to the extent that the leader employs an array of four strategies to direct the discussion For each segment of the text discussed, the leader frames a question to which the group members respond The participants are then free to share additional questions they generated while reading the material The leader summarizes by identifying the gist of the segment that has been read The group then comments upon the summary offered by the leader At any time, in the reading or discussion, the group is encouraged to note when content in the text or a point raised in the discussion should be clarified Finally, the leader signals preparation to move onto the next portion of the text by making and soliciting predictions regarding upcoming content. In summary, reciprocal teaching is an interactive teaching procedure in which the teachers and students collaborate in the joint construction of the meaning of text

Underlying the model of reciprocal teaching is the theory that expert led social interactions have a prominent role to play in learning and can provide a major impetus to cognitive growth While this idea is most closely identified with Vygotsky (1978), a number of theorists, including

Binet (Binet, 1909, Brown, 1985), Dewey (1910), and Piaget (Piaget, 1967, Brown & Reeves, in press) emphasized guided learning in social contexts as a key to developmental change

Previous investigations of reciprocal teaching supported this model of instruction (Brown & Palincsar, in press; Palincsar & Brown, 1984; Palincsar, 1986) The transcripts of the intervention revealed that discussion presented an excellent tool for the instruction of these cognitive strategies; the teachers had ample opportunity to model use of these strategies when they led the discussion as well as opportunities to guide the students in their use of the strategies when the students led the discussion The improvements in the students' abilities to execute the strategies (i e., to write a summary, to generate a set of teacher-like questions and to detect incongruous lines in text) suggested that what had been learned in the group context was, in fact, internalized by the students Furthermore, significant improvements in the accuracy with which the students completed sets of comprehension questions, measuring the recall of text-explicit, text-implicit, and script-implicit information (administered each day of the intervention and using assessment materials independent from the training materials), suggested that the instruction did indeed promote increased comprehension Finally, and perhaps most significantly, gains that were demonstrated in the experimental setting were observed to generalize to other classroom settings.

The study which I will report today was designed as a replication of our initial reciprocal teaching interventions. However, there were significant differences between this work and former studies First, the

six teachers participating in this study were not volunteers, and three of them were not reading specialists. Their groups were more heterogenous with regard to both decoding and comprehension ability than our former groups. Finally, the groups were much larger than those with which we had previously worked (which ranged from four to eight).

Collaboration with the teachers

Let me begin by describing the teachers with whom we collaborated. The six teachers had each taught a minimum of ten years, three of the teachers were trained as reading teachers while the other three had teaching endorsements but backgrounds in theology, library science. The study we proposed to these teachers was rather intrusive and extensive. We were asking these teachers to suspend their current reading programs in one of their classes and to tolerate our presence in two of their classes for over thirty school days, which, as those of you who have conducted research in the schools will attest, can easily take three months. To determine just how intrusive this intervention would be, we asked the teachers to complete a survey regarding their current instructional practices and conceptions regarding reading and the pedagogy of reading. To summarize, these six teachers indicated that they relied heavily on published remedial materials that taught their students basic skills such as using structural analysis and context cues to decode and understand unfamiliar words. The teachers used individualized instruction as necessary but generally instructed on a whole-group basis, dividing each period into oral reading, silent reading, and worksheet completion. Generally speaking, the teachers attributed the comprehension

problems experienced by their students to lack of interest or motivation (f=6), poor attention span (f=4), inadequate vocabulary knowledge (f=4), and word identification problems (f=1)

Frankly, we didn't share a lot of common ground, in terms of a number of instructional goals, teaching practices, or conceptions about reading and poor readers. In fact, as the teachers and I discussed the survey and contrasted their program with the intervention that was targeted in this particular study I understood, from the teachers' perspectives why, in some dictionaries the phrase "see treason" accompanies the definition of collaboration Nevertheless, it was clear that we did share an agenda; the teachers identified their number one aim as improving the comprehension skills of their readers regardless of their ability level This was our shared goal

Teacher Preparation

At this point in the study, we worked more as consultants with the teachers than as collaborators We attempted to employ the same model of scaffolded instruction in the preparation of the teachers that would be employed when working with the students During the initial in-service, the teachers reflected on their current remedial programs in terms of goals and instructional activities and completed the survey discussed above The investigator discussed current research regarding the features of successful reading comprehension instruction, focusing particularly on the role of self-regulatory strategies in reading Reciprocal teaching was introduced by describing the theory which informed its design The following points

were emphasized (1) The acquisition of the reading strategies employed in reciprocal teaching would be a joint responsibility shared by the teacher and students; (2) The teacher would initially assume the major responsibility for the instruction of these strategies by modeling, in a very explicit fashion, the process of using these strategies (i.e., the teachers would "think aloud" how they generated a summary, what cues they used to make their predictions, how they had used rereading or reading ahead when they encountered something unclear in the text); (3) All students would be expected to participate in this discussion; that is, all students were to be given the opportunity to assume the role of teacher and lead the discussion To ensure that the students participated successfully, the adult teachers would enable the students' participation by prompting the student or by altering the demand For example, if a student were unable to summarize when it was his or her turn, the teacher might ask, "what was the topic of the part we just read?," "What did you learn about that topic?" and then build upon the student's response to generate a summary; (4) Throughout each day of instruction, there would be a conscious attempt to release control of the dialogue to the students, to fade the leadership of the teacher so that instead of modeling and instructing, the teacher would be providing feedback We have compared this role to that of a coach and used the same metaphor with the teachers The teachers then viewed a videotape of a reciprocal teaching lesson being conducted with a group of eight junior-high students

In addition to this day of in-service training, there were two more training sessions conducted in the teachers' schools (Three teachers

worked at one school, two teachers worked at a second school and one teacher worked at a third) At the second session, the teacher(s) and investigator role-played the intervention using the instructional passages The investigator first role-played the teacher and then role-played the students, simulating the situations that had arisen in previous research (e g , summarizing by reading virtually intact paragraphs) Appropriate responses were discussed The final formal training session was again held in each of the schools During this session, the investigator and teacher(s) worked with a group of students who would not be involved in the research but who were similar to the experimental students. The process of introducing the instruction, structuring the discussion with the use of the four strategies, and guiding students' participation in the dialogue was modeled and discussed, with the investigator and teachers taking turns leading the instruction While this marked the last structured training session, additional coaching was provided each teacher on a weekly basis throughout the intervention The extent of this coaching, which was more in the nature of collaboration, will be discussed later in this paper Furthermore, the teachers met once, midway through the intervention, to share their observations regarding the intervention.

The student collaborators

Student Selection

We identified the students by screening each of the six developmental reading classes instructed by the six middle-school remedial reading teachers The screening proceeded by having the students read silently a 500

word passage written at the seventh grade level entitled, Paper: A Web for Words. They were cautioned to read carefully so that they would be able to answer comprehension questions about the story. The students were also told to ask for assistance with any word(s) they could not read or understand. Upon completing the reading, the students received ten questions to respond to in writing, from recall of the story. To determine correct and incorrect oral reading rates, the investigators then listened to each student read aloud a 175 word sample from the same passage which had been read silently.

For each of the six teachers, the two classes containing the largest number of students who met our criteria for demonstrating a discrepency between decoding and comprehension ability were identified. Our criteria included the ability to read grade level material with 80 word-per-minute accuracy and two or fewer error words-per minute (Streng), while attaining 50% or less correct on the comprehension questions. In only one instance did a teacher have more than two groups in which a vast majority of the students attained these criteria. In this instance, we selected the two groups of equivalent size. The two groups per teacher were then randomly assigned to the treatment or control conditions.

Students

The experimental groups totaled 71 students and the control groups included 76 students. The data for any student not present for at least 10 days of instruction/assessment were not included in the final analyses of the data. By that standard, 63 students in the experimental group and 66 students in the control group produced data that were analyzed.

Table 1 describes the experimental and control groups in terms of their grouping by teacher, assignment to experimental or control conditions, number of students in the group, pretest scores on the Gates-MacGinitie comprehension and vocabulary subtests, correct and incorrect oral reading rates, and accuracy on criterion-referenced measures of comprehension. The smallest experimental group numbered seven (Teacher 4) while the largest numbered 16 (Teachers 2 and 6) Control group sizes ranged from nine (Teachers 1, 3, and 5) to 19 (Teacher 6) For four teachers, the experimental and control groups were within three children of being equivalent in size. For Teacher Four, the control group exceeded the experimental group by seven and for Teacher One, the experimental group exceeded the control group by six

Comprehension, as measured by the G MacGinitie comprehension subtest, ranged from about three and a half years below grade level (for the Experimental and Control students for Teacher Three) to about one half year below grade level (for the experimental group for Teacher Four and the control group for teacher five). Comprehension scores, as measured by five criterion referenced assessments administered before intervention, were less variable within teacher, and ranged from 34% for Teacher Three's experimental group to 60% for Teacher Two's control group

Thirty-three percent of the students were minority students, the majority of these students being black Minority students were represented fairly evenly across the instructional groups. A small percentage of minority tudents were from Southeast Asia; however, all had achieved fluency with English With few exceptions, the students in each of these groups had n

history of receiving services in Chapter 1 reading programs for at least two years prior to the year of this study

Pre-testing

Prior to the intervention, the investigators administered the pretest on the transfer measures with the experimental and control classes in the following manner All groups were told that the investigators were interested in how junior high students remember and understand what they read and how the activities they do in reading class affect their ability to read The summarizing task was then introduced by eliciting from the class a definition of a summary, emphasizing that a summary is a shortened version containing the most important information Rules for constructing summaries were then written on the board and illustrated Taken from the work of Brown and Day (1983) and Kintsch and VanDijk (1978), they were: (1) decide if there is a topic sentence; (2) make up a topic sentence if one is not provided in the text; (3) do not include what is unimportant; (4) do not repeat information; and (5) name lists The students were given two pieces of paper for each summary On the first, they were asked to write a draft of their summary The second paper contained 60 lines The students were asked to reduce their drafts to 60 words, writing one word on each line The students completed two summaries during pretesting over two days

The question-generating pretest was administered by giving the students a passage and requesting that they write ten questions a teacher might ask if she were testing them on the material in the passage They were told not

to ask true/false or fill-in-the-blank questions. The students completed the question-generating task for two passages during the pretest over two days

Also completed prior to the intervention, were five criterion-referenced comprehension assessment passages. These were administered in the same manner as the screening passage. The students read each passage silently, asking for assistance with unfamiliar vocabulary, and answered the ten accompanying questions from recall. Throughout the study, the students in both experimental and control classes were shown a graph depicting the results of their assessments on a weekly basis. The teachers distributed the graphs to each student. They explained to the class as a whole what the axes stood for and what the plots indicated. They then walked around the room, answering questions and commenting on students' graphs. Completion of the pretests and baseline measures of comprehension required, on the average, five class periods

Intervention

Experimental Group On the first day of intervention, the teachers engaged their experimental classes in a discussion regarding why it is sometimes difficult to understand and remember what they have read The teachers then explained that for the coming weeks, the students would be learning four activities to help improve their reading comprehension ability. The four strategies were described, along with the way in which they would be taught (the discussion format and turn taking in the role of teacher) as well as how progress in learning those activities would

be measured and shared (daily assessments and weekly graphs). The final point of discussion was the variety of situations beyond developmental reading class in which it would be appropriate to engage these activities (i e., to prepare for tests, for classroom discussions, for writing book reports, etc). The manner in which the students would be introduced to the procedure was fairly well-scripted for the teachers

The students then completed, with teacher guidance, a series of worksheets designed to introduce the four strategies. These worksheets, which presented a new feature of the intervention, were constructed to ensure that the students possessed minimal competency regarding the four strategies before the reciprocal teaching dialogue commenced (i.e., that all students could use a question form, could recognize that titles and headings can frequently be used to predict content, and that a summary captured main idea (information) To illustrate the format of these worksheets, we will describe how the students were introduced to question generating After a brief discussion about the role that questions play in our lives, particularly our school lives, the students were asked to generate information-seeking questions about everyday events This activity permitted the teachers to ascertain how well their students could frame a question The students then read simple informational sentences about which they were to ask a question and were supplied words they might use to begin their questions The question words faded out and the sentences grew longer Next, The students were given a paragraph and three accompanying questions they had to evaluate One of the questions could not be answered from the text, the second concerned a trivial fact, and the third represented

a main idea question. Finally, the students had to generate their own questions to accompany the paragraphs. Following these worksheet activities, which took five days to complete, the students and teachers undertook the reciprocal teaching procedure described earlier in this paper. A more specific description of its implementation in this study follows. The program consisted of 20 days of reciprocal teaching instruction. Each session lasted approximately 40 minutes and was followed by the administration of an assessment passage.

Implementing reciprocal teaching

For the first five days, the teachers began the lesson by reviewing the purpose of the reciprocal teaching instruction. The strategies the students would use to structure the discussion were elicited from and defined by the students. The teachers then distributed the passage to be used for the day. If it were a new passage, the group began by making predictions about the content based upon prior knowledge regarding this topic or based upon information they would hope to acquire from reading the passage. If it were a passage that had been introduced previously, the teacher began by eliciting a summary from the students. The introductory discussion was then concluded with a series of predictions.

Throughout the intervention the adult teacher assumed responsibility for the first segment of text to be read. The group would read silently a designated portion of the text. For the first ten days of instruction, these were generally paragraphs; the length of segments were then increased to cover portions of text, for example, from one subheading to

the next The adult teacher would lead the discussion by posing a question she considered relevant to the text The students would answer the question and the teacher would ask if there were additional questions the students had thought of as they read the segment The teacher would then summarize, sharing with the students how she had used the summarization rules to assist her with her summary The teacher would ask if the students would have modified her summary The teacher then noted anything in the segment which seemed unclear upon initial reading or about which she still had questions If the former were the case, the teacher discussed how she eventually came to an understanding of the content If it were the latter, the teacher would frame these questions as her predictions about upcoming text The teacher again invited the students to add any clarifications they had made while reading or indicate any points which still required clarification

Before moving to the next segment of text, a new discussion leader would be assigned This assignment, as well as subsequent assignments, were made on a voluntear basis, with the provision that everyone would have a turn as discussion leader each day or every other day (depending upon group size) The teachers maintained a record of student involvement by placing a check next to each student's name when they entered the discussion and/or took the turn of teacher This was particularly useful for those classes where there were more than ten students Following the assigned silent reading, the discussion leader would lead the group through the discussion of the text. The adult teachers would encourage the discussion necessary to achieve concensus regarding the important questions to be derived from the text as

well as the summary of the text While clarifications and predictions were invited with each segment, the text would not always support a prediction or suggest a clarification. The students were asked to make this evaluation Generally, the adult teacher would then comment upon the performance of the student teacher and ask the student teacher to select the next teacher Following forty minutes of instruction, the adult teacher would give general feedback to the group and invite the group to compare the session to yesterday's or to comment upon which of the strategies they were learning particularly well. She would then encourage the students to remember these activities as they independantly read the day's assessment. The students were then given a copy of their assessment passage for the day. As they completed reading the passage, they would raise their hands, and the classroom teacher would take back the passage and give them the accompanying questions.

Control Groups

The students in the control group continued to work in the curriculum adopted by the developmental-reading teachers. In general, these students read from basal readers, completed accompanying workbook activities, and supplementary worksheet packets planned for each student Generally speaking, the students read orally for fifteen minutes, then were assigned a story for silent reading which required fifteen minutes, and then completed the accompanying worksheet activities for ten minutes The teachers then corrected these with the students individually In addition, the control students completed the criterion-referenced comprehension

assessments in the same manner as the experimental students. During pretesting the control students received the same explanation regarding the purpose of the study as the experimental students; to determine how the activities they did in reading class affected their ability to understand and remember what they read.

Maintenance and Posttesting

At the completion of the 20 days of instruction, on the next school day, the students entered a five day maintenance phase during which they completed one assessment per day. Following the last maintenance check, the students in both the experimental and control classes completed the summarization and questioning transfer measures as a posttest.

Procedural Reliability Checks

An investigator attended each of the experimental and control classes two times a week minimally. The purpose of these observations was to collaborate with the teachers in implementing the experimental procedure and to monitor the instruction received by the control groups. In addition to these observations, each reciprocal teaching lesson was audiotaped.

RESULTS

There are several sets of results to be considered. The first is an evaluation of the reciprocal teaching sessions themselves. The second will be the results of the daily criterion-referenced assessments of comprehension and the third will be the transfer measures of summarization and question generating. Finally, assorted anecdotal observations will be shared.

The reciprocal teaching sessions

Of the six groups, there were three who appeared, virtually from the outset, to be very comfortable with the procedure In fact, during the training sessions, when we role-played the use of the procedure as a group, other teachers commented upon the ease with which the teachers of these three groups worked. Collaboration with these teachers was a matter of sharing impressions regarding the sessions and discussing the progress of individual students in the sessions as well as on the daily assessments.

We collaborated more actively with the remaining three teachers The teacher of group three indicated, and we agreed, that this discussion was a real struggle for her group She confided that she felt it was probably due to her own feelings of inadequacy with this type of instruction combined with the fact that she was not, by preparation, a reading teacher While all of the above might have been true, it was also the case that she had the poorest of our six experimental groups as indicated by four of our measures (standardized comprehension test, correct and incorrect reading rates, and baseline assessment) Perhaps to compensate for their low achievement this teacher was very directive and mechanical in her use of the procedure. For the initial three days, the "discussion" took on the form of teacher inquiry and student response with no indication that this pattern would change We collaborated with this teacher by participating in the groups from days four through eight, modeling how the instruction could be restructured to reflect discussion; for example, asking the students to comment on the summary generated by the leader and having the students indicate whether their question had been answered correctly

We also collaborated more actively with groups one and five. Group one had fifteen children in it and group five had only nine; however, eight were boisterous boys. After observing for three days, the objectives with these groups was just the opposite of the objectives described for the previous teacher. Our goal was to have these teachers begin implementation of reciprocal teaching in a more structured fashion. Both of these teachers had attempted to very quickly place the students in control of the discussion, however, the students hadn't indicated readiness to do so and, consequently, the "discussion" quickly deteriorated into something not unlike a verbal "free-for-all." We joined these groups for the next six days, modeling a more structured use of dialogue where the teacher maintained a more active role in leading the discussion. We also modeled and discussed a structured way in which the teachers could relinquish control of the discussion. For example, we encouraged these teachers to begin each session by taking their turn as discussion leader first, then call upon a better student for the next turn, lead the discussion once again themselves, and so on. We observed and coached after the sessions from days ten through twelve and merely provided feedback through days 15.

To validate our own observations of these sessions, we asked two raters, both of whom had conducted reciprocal teaching instruction, to listen independently to sessions from the first three and the last three days of instruction for each teacher and to rank order the quality of instruction. For the first half of instruction, while the raters were not in 100% agreement regarding the exact ordering of the lessons, they did rate the three teachers who worked independently from the beginning among

tha top thraa and concurrad on tha two bottom ranked laasons For the aacond half of inatruction, they concurrad on tha top two laasons but were not conaiatant in their ratinga of the other four groupa. The ratara noted that aach of the taachara indicatad fidelity to the raciprocal taaching procedure.

In aummary, the quality of inatruction, whi'a aomawhat variabla, was controllad by the aaaiatanca that waa provided to the taachara. Thraa of the aix taachara raquirad aupport aftar the formal training waa diacontinuad. Tha practica of acaffolding for the taachara through the uae of modaling, aalf-appraiaal and faadback, additional inatruction, and ancouragamant waa an affactiva maana of aaaiating the taachara to achiava compatanca. Analyaaa of tha dependant maaaura uaad to aaaaaa tha affactivanaaa of the intarvention will now be praaantad.

Comparability of the experimental and control groups prior to inatruction

To datarmina the comparability batwaan tha axparimantal and control groupa prior to intarvention, a multivariata analyaia of variana for rapaatad maaaura waa uaad. MANOVA waa aalactad baceuaa taata baaad upon thia approach ara fraa of apharicity aaaumptiona (O'Brian & Kaiaar, 1985). Tha four dependant maaaura were the mean for the fiva aaaaaament paaaagaa adminiatarad in baaaline, tha acoraa aarnad on the prataat aummary and quaation ganarating taaka, and the correct decoding rata of aach atudant. Tha maana and atandard deviationa for thaaa maaaura ara raportad in Table 1. Tha raaulta indicatad that thara waa no aignificant diffarana batwaan tha axparimantal and control group nor waa thara a aignificant intaraction

between group and teacher. There was however, a significant overall teacher effect, F (20,414) = 2 04,p < 005 A series of Bonferroni T tests indicated that this effect was true only for the means on the baseline comprehension assessments and not for the other dependent measures On these assessments, there was a significant difference between those classes taught by teachers two and three, two and five, six and three and six and five

Performance on the criterion-referenced comprehension measures

To determine the effect of the intervention on students' responses to the comprehension assessments, MANOVAs were conducted comparing the change scores across phases That is, the phase mean for baseline was compared to the mean for the first ten days of instruction; the mean for the second ten days of instruction was compared to the mean for the first ten days, and the mean during the maintenance phase was compared to the mean for the second ten days of instruction The means and standard deviations for these analyses are presented in Table 3 Examining the effect of the intervention over time, the MANOVA revealed a significant effect for group, F(1,5) = 8 97, p < 05, indicating that, in comparison to control children, the experimental children were more accurate in their answers to the comprehension questions overall In addition, there was a significant effect for teacher, F(5, 117) = 6 79, p < 001, which indicated that some teachers' classes performed better than others. There was also a significant effect for time of testing, F (3, 115) = 105 86, p < 001, a significant time of testing X group interaction, F (15, 317 87) = 3 72, p

< 001 and a significant time of testing X group X teacher interaction, F (15, 318 87) = 3 37, p < 001

To examine the effects of each phase of the intervention, a series of contrasts were computed, comparing the means from one phase to the next The first analysis compared performance on baseline with performance during the first half of intervention This analysis yielded a significant main effect for teacher F(5, 117) = 2 97, p < .05 and a significant group X teacher interaction, F (5, 117) = 5.36, p < 001, indicating that teachers' groups improved differentially from baseline to the first half of intervention The average score of the experimental students increased by 20 13% as compared to a 12.01% increase on the part of the control students.

The next contrast examined whether the second half of the intervention had an effect over and above the effect of the first half of intervention This contrast yielded a significant group effect F (1,5) = 6 69, p < 05 The average score for the experimental group went from 64 95% to 72 0 whereas the average score for the control children actually decreased from 59 55% to 54 27%

A final contrast, comparing the means for the second half of intervention with the means for maintenance, resulted in no significant effect (F < 1) This indicates that students maintained their level of accuracy with the assessments following the intervention

To pursue further the time of testing X group X teacher interaction, separate univariate F tests and Bonferroni t tests were conducted on each phase of the intervention Differences during baseline have already been

reported An ANOVA on the mean comprehension scores for the first half of intervention revealed a significant affect for teacher, F (5, 117) = 7 44, p < 001 Bonferroni tests revealed a significant difference between the experimental and control groups in teacher 6's class (means: E = 75 92, C = 57 24) In addition, the experimental groups of teachers 2 and 6 (means E2 = 75 4 E6 = 75 92) performed significantly better than the experimental and control groups of teacher 3 (means E = 48 14, C = 46 67) and the experimental group of teacher 5 (mean = 53 88) Teacher 4's experimental group (mean = 76 0) also performed significantly better than teacher 3's experimental and control groups

An ANOVA on the comprehension scores for the second half of intervention revealed a significant main affect for group F (1, 5) = 32 50, p < 01, indicating that, overall, experimental groups outperformed control groups There was also a signficant effect for teacher, F (5, 117) = 4 7, p < 001, indicating that some classes performed signficantly better than others However, Bonferroni tests failed to reveal any significant comparisons between teachers Therefore, although the teachers' classes were scattered in such a way that there was an overall significant affect for teacher, no two pairs of teachers were significantly different to each other

The difference between the experimental and control groups was maintained after instruction was discontinued The ANOVA on the means for maintenance revealed a significant main affect for group, F (1, 5) = 18 03, p < 01, indicating that overall, children in the experimental groups performed significantly better than children in the control groups A

significant main effect for teacher, $F(5, 117) = 5.42$, $p < .001$ indicated that some teachers' classes performed better than others. The results of the Bonferroni tests revealed that, overall, teacher 2's class performed significantly better than teachers 1 and 3's classes.

To summarize, there was a significant improvement in experimental students' accuracy with the recall and interpretation of material they read independently as the days of instruction proceeded. Furthermore, the majority of the instructional groups indicated this improvement after the first ten days of instruction. Improvement continued throughout the second half of intervention and was maintained following completion of instruction. To render the assessment data more useful, it is helpful to think of the proportion of students in the experimental and control groups who achieved criterion performance. Based on the performance of normal readers completing these assessment passages, we designated criterion performance as 70% or greater accuracy on four out of five consecutive days. In the experimental group, greater than 70% of the students attained criterion performance, whereas only 25% of the control students attained criterion performance. Furthermore, of the experimental students who did not achieve criterion performance, another 26% indicated at least a 30 point gain over baseline that maintained following the intervention. Only 3% of the experimental students failed to indicate any gains as compared to 60% of the control students.

While it would be intriguing to explain teacher effects, any interpretation would almost certainly be too simplistic. One would have to disambiguate the effects of entry level performance on the dependent measures, class

size, and heterogeneity of the groups across all measures including decoding ability. For example, in the case of teacher 6, who had the largest group of students, her experimental group achieved the greatest degree of accuracy; however they were also the highest group to begin with on the assessment measure as well as one of the most homogenous groups. Almost certainly the within-child as well as within-class variables interact with teacher effectiveness to yield the results that were obtained.

The transfer measures

There were two transfer tasks, summarizing and question generating. Although these tasks constitute two of the four strategies employed in the intervention, they represent transfer measures to the extent that they were assessed as a writing activity and the students needed to consider an entire piece of text in the process of completing these measures, as opposed to summarizing or generating questions for portions of text.

The first transfer task to be considered is summarizing. Independent raters scored the students' summary sheets and assigned points in the following manner: One point was awarded for each superordinate used. A point was assigned for each topic sentence provided by the text that was incorporated in the summary. Students were given two points if they invented a topic sentence. Every proposition included in the summary was assigned a point if the corresponding proposition had been rated by a group of experts (English teachers) as quite to very important. Ideas rated as unimportant by the raters and included in the students' summaries were not assigned points. Beyond these points, points were assigned to reflect

overall clarity and quality of the summaries. The number of points assigned to the two summaries completed before and the two completed after the intervention were tallied and used in the analysis. The raters were unaware of the condition in which each student had participated. Furthermore, they were unaware whether the summary was completed during pretesting or posttesting. In addition to rating their own sets of summaries, both raters scored 25% of the others' summaries to determine inter-scorer reliability. The Pearson R was .98. (Incredible, I know.)

The MANOVA indicated a significant group by treatment interaction, F (1,5) = 6.84, $p < .04$. A followup ANOVA indicated that there was a signficent difference between the experimental (mean = 12.15) and control groups (mean = 14.74) on the pretest summary score with the control group scoring signficantly higher than the experimental group F (1, 127) = 10.52, $p < .002$. Therefore, ANCOVA was used to examine the posttest scores of the experimental and control groups, using the pretest summary score as the covariate. The ANCOVA resulted in a significant difference between the experimental and control groups. The mean score attained by the experimental group on the posttest (16.38) was significantly higher than that achieved by the control group (13.82), F (1, 127) = 12.69, $p < .001$. An examination of the points earned by the experimental and control groups indicated that the experimental students' summaries differed most on the number of invented topic sentences employed in their posttest summaries as well as in their use of superordinates.

Turning to the question-generating transfer measures, these results were scored by the same two independent raters who rated each question as:

a main idea question (worth two points) or a detail question (one point), as a question lifted directly from the text (0 points) or paraphrased (one point). In addition, the quality of the question was rated on a five-point scale from one (very poor) to five (excellent). Finally, a question which the rater indicated she would ask if she were testing children on the passage was awarded an extra point. The results of this measure were not consistent with the summary transfer measures. The pretest mean for the experimental group was 44.9 (S.D.= 13.8) while the mean for the control group was 45.2 (S.D.= 14.8). The posttest mean for the experimental group was 50.8 (S.D. 12.6) while the mean for the control group was 50.3 (S.D.= 14.8). These means represent a significant overall differences between the pre- and posttests, $F(1, 123) = 24.68$, $P < .0001$. There were no significant overall group-by-treatment interaction. Furthermore, while there was an overall significant teacher effect, $F(5, 118) = 4.97$, $p < .0004$, there was no overall significant teacher-by-group by-treatment interaction. There are at least two conceivable explanations for these results; one is that both the experimental and control groups received considerable practice answering questions during the course of the intervention. Perhaps the repeated exposure to questions enhanced their ability to generate questions independently. Furthermore, a number of teachers attended a workshop at which the game "hot seat" was introduced. In this game, the students read a story, one student is then assigned to sit on the hot seat while the rest of the class quizzes that student. When the student fails to answer a question correctly, the person who asked the stumping question, assumes the hotseat. Four of the six teachers implemented the hot seat game during the

intervention with their control students.

To further understand the relationship among the dependent variables, a series of multiple regressions were performed. To summarize the most interesting of these analyses: a significant amount of variation on the baseline comprehension assessments is explained by pretest performance on the transfer measures (multiple $R^2 = .32$, $p < .0000$); improvement on the comprehension assessments can be predicted from improvement on the summary transfer measure, $F(3, 114) = 2.79$, $p < .02$; while decoding ability is a significant predictor of baseline performance and performance during the first half of intervention on the comprehension assessments, it ceases to be a significant predictor of performance on the comprehension assessments administered during the second half of intervention and during maintenance

Anecdotal data

Final indices of the effects of the reciprocal teaching intervention are anecdotal data. These data can be described as teacher-related and student-related. When the teachers completed the same survey they were asked to complete prior to the study, there were interesting conceptual changes indicated by the teachers. For example, when asked to describe why students have difficulty understanding what they read the responses now included: "students must learn skills that lead to better understanding," "students must be taught that comprehension is no accident," "students must not lose sight of the reason why they are reading - to understand- they cannot be so intent on word-by-word reading." Of the six teachers involved in this study, four continued to use the intervention as described

the next year, the teacher from group three who indicated discomfort with the procedure used the intervention as a writing activity (Study 2) The remaining teacher used only the assessment procedure the following year. Of the six teachers, two have given state-wide in-services regarding reciprocal teaching and one, who also teaches adult literacy classes, has conducted reciprocal teaching with these students as well

It is unfortunate that we administered no measures of the students' attributions, perceived self-competence, or metacognitive knowledge. We have anecdotal indications such as students in the experimental group identifying that the assessments contained different types of questions, or that there were certain passages that were going to be more difficult for them because they knew less about the topic. We know that, for the most part, the students enjoyed this activity In fact, teachers who were initially concerned about behavior management, reported that behavior management was the least of their problems during the intervention and attributed this to the interactive nature of the instruction as well the seriousness with which the students took the opportunity to assume the role of the teacher

SUMMARY

This is a paper about several issues. First, it is of course, a report of research on strategy instruction conducted by classroom teachers with their remedial-reading students Several measures that approximate the demands students encounter in school life, indicate that it was a successful effort at cognitive instruction. This is good, but it's not news

To appreciate what is "news-worthy" about this research is to go behind

the scenes. It is to recall that, for good or for ill, prior to this study, none of the six teachers in this study conducted instruction that looked at all like the intervention, nor did any of these teachers express a conception of reading or reading instruction that matched that held by the investigators. It is to think about the fact that only three teachers, after fairly extensive inservice work (extensive as least as compared to traditional in-service models) were prepared to conduct the intervention as it had been designed. And it is to recognize that by using the same principles of scaffolded instruction with the teachers as we were advocating be used with the students, the teachers became far more competent.

While the quantitative data are important to this study, the data that are more informing include: the dialogues, providing a glimpse as to the dynamic nature of scaffolded, prolaptic-instruction, the opportunities for collaborative problem-solving as well as for individualized diagnostic teaching that emerge in this instruction. Outcomes as important as the improvement in students' abilities include indicators of conceptual change on the part of the teachers and maintenance, not only on the part of the students, but also on the part of the teachers who added reciprocal teaching to their repertoire of instructional strategies and redefined some of their instructional objectives.

Notes

1

Materials

The passages read during instruction, the criterion-referenced comprehension assessments, and the transfer test materials employed in this study were identical to those used in earlier work (cf Palincsar, 1982; Palincsar and Brown, 1984) A total of ten passages, averaging 1500 words each, were used during the intervention with the experimental groups These expository passages, determined to be written at a seventh grade level according to the Fry Readability Formula, were selected from the following reading series: Reading Unlimited (Aaron et al , 1976); Keys to Reading (Matteoni, Lane, Ducher & Burns, 1980); Adventures for Readers (Nieman & Safier, 1979); Reading 720 (Clymer, 1976); Corrective Reading Decoding (Engleman, Becker, Hanner & Johnson, 1978); and Serendipity (Durr, Pescosolido, & Toetter, 1974) These passages covered such topics as the uses of solar energy, the mining of salt, and the sport of falconry All teachers presented these passages in the same sequence

2

The criterion-referenced assessments

The C-R assessment passages were taken from the same sources as the instructional materials; however, they were adapted so that they were shorter in length (typically 475 words each) and would lend themselves to the construction of ten comprehension questions. These ten questions were developed using Pearson and Johnson's (1978) taxonomy Four questions were text explicit, the answer was explicitly provided in the text. Four questions were text implicit, the answer needed to be inferred

by integrating information presented in the text; and two questions were script implicit, the answer was achieved by considering the information in the text in relation to one's prior knowledge of the topic at hand. Each set of questions was rated independently by two English teachers on two features. They first indicated the category of question (T.E., T.I., or S.I.) and then the difficulty level (1=easy, 2=moderate, 3=difficult) of each question. Questions were rewritten until there was 100% agreement on question type and the level of difficulty across the sets of questions was fairly consistent. The order in which these assessment passages were administered was randomized across all students in the study.

3
The transfer tasks

The materials used in the transfer tasks were of two types. There were four 500 word stories to be summarized. These stories were employed in the summarization work of Day (1980) and were specifically constructed to facilitate use of the summarization rules to be described later. In addition, there were four 500 word passages that were used to elicit 10 student-generated questions each. The transfer passages were counterbalanced for order of presentation, two were administered prior to the intervention and two were administered following the intervention.

4
Results not available

Results that will not be considered include an aborted attempt at conducting generalization probes in the science classes. One probe was conducted during baseline and one during the first half of intervention

Unfortunately, absenteeism was a significant problem during both probes furthermore, the students who were absent during the baseline probe were not the same ones absent during the intervention probe There was a certain degree of teacher resistance to the probes and the data we were able to derive were not worth pushing our luck in the system

References

Binet, A (1909). Les idees modernes sur les enfants. Paris: Ernest Flammarion

Brown, A L. (1985) Mental orthopedics: A conversation with Alfred Binet In S Chipman, J. Segal, & R Glaser (Eds.), Thinking and learning skills Research and open questions (Vol. 2, pp. 319-337). Hilladale, NJ: Erlbaum

Brown, A & Palincsar, A S (In press). Guided, cooperative learning and individual knowledge acquisition In L.B Resnick (Ed.) Cognition and Instruction: Issues and agendas Hilladale, NJ: Lawrence Erlbaum Associates.

Brown, A L & Day, J.D (1983) Macrorules for summarizing texts: The development of expertise Journal of Verbal Behavior, 22, 1 14

Dewey, J (1910) How we think Boston: Heath. (2nd ed , Heath, 1933).

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

O'Brien, R G & Kaiser, M K. (1985). MANOVA Method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 97, 316-333

Palincsar, A.S. (1986) The role of dialogue in providing scaffolded instruction. Educational Psychologist, 21 (1 & 2), 73-98.

Palincsar, A S & Brown, A L. (1984) Reciprocal teaching of comprehension fostering and monitoring activities Cognition and instruction, 1 (2), 117-175

Piaget, J. (1967) The grasp of consciousness: Action and concept in the young child Cambridge, MA: Harverd University Press.

Slavin, R. E (1983) Cooperative learning. New York: Longman, Inc

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

TABLE 2

TEACHER-GROUP	BASELINE	INTERVENTION Phase 1	INTERVENTION Phase 2	MAINTENANCE
1-Experimental	45.9	60.3	65.5	63.6
	SD 19.6	SD 15.4	SD 13.9	SD 15.4
Control	44.9	56.3	56.3	46.0
	SD 23.0	SD 19.9	SD 18.2	SD 27.0
2-Experimental	51.0	75.4	74.5	77.0
	SD 11.6	SD 8.1	SD 6.9	SD 9.6
Control	59.9	67.7	60.7	72.3
	SD 14.6	SD 15.0	SD 15.7	SD 13.2
3-Experimental	34.0	48.1	61.3	68.1
	SD 11.1	SD 7.2	SD 13.7	SD 10.0
Control	41.6	46.7	53.7	51.1
	SD 14.7	SD 16.2	SD 15.7	SD 17.6
4-Experimental	45.7	76.0	80.7	82.3
	SD 16.9	SD 11.6	SD 2.7	SD 8.1
Control	45.2	53.8	49.0	56.5
	SD 15.0	SD 20.3	SD 24.7	SD 23.2
5-Experimental	35.5	53.9	65.5	65.4
	SD 9.2	SD 7.7	SD 9.2	SD 16.2
Control	43.1	56.3	51.1	54.3
	SD 16.8	SD 13.6	SD 19.6	SD 7.9
6-Experimental	56.8	75.9	83.4	84.0
	SD 14.0	SD 8.5	SD 7.1	SD 4.9
Control	53.6	58.6	65.8	58.6
	SD 11.1	SD 12.5	SD 14.6	SD 16.2

TABLE 1

Descriptive Statistics for Students in Study 1

Teacher	Group	n	Comprehension		Vocabulary		Oral Reading		Percent Correct:
			GE	Delay	GE	Delay	Correct Rate	Incorrect Rate	Baseline Assessments
1	E	15	5.5	-1.5	5.3	-2.3	112.7	2.2	45.9
1	C	9	4.9	-2.1	4.3	-2.7	80.2	1.7	44.9
2	E	16	5.4	-1.6	6.3	-1.3	117.1	1.8	51.3
2	C	16	5.3	-1.7	5.6	-1.4	120.5	0.9	59.9
3	E	8	3.6	-3.4	5.9	-1.1	103.2	2.9	34.0
3	C	9	3.3	-3.7	4.3	-2.7	108.9	2.7	41.5
4	E	7	6.4	-2.6	6.5	-2.4	126.9	0.9	45.7
4	C	14	4.7	-3.3	5.5	-2.5	125.3	1.3	45.2
5	E	9	4.3	-2.7	4.7	-2.3	119.3	1.7	35.5
5	C	9	6.6	-0.4	6.3	-1.3	112.1	1.8	43.1
6	E	16	5.0	-2.0	5.8	-1.2	118.5	1.8	56.8
6	C	19	4.6	-2.4	5.6	-1.4	100.7	2.3	53.6

GE = Grade Equivalent - Gates-MacGinitie

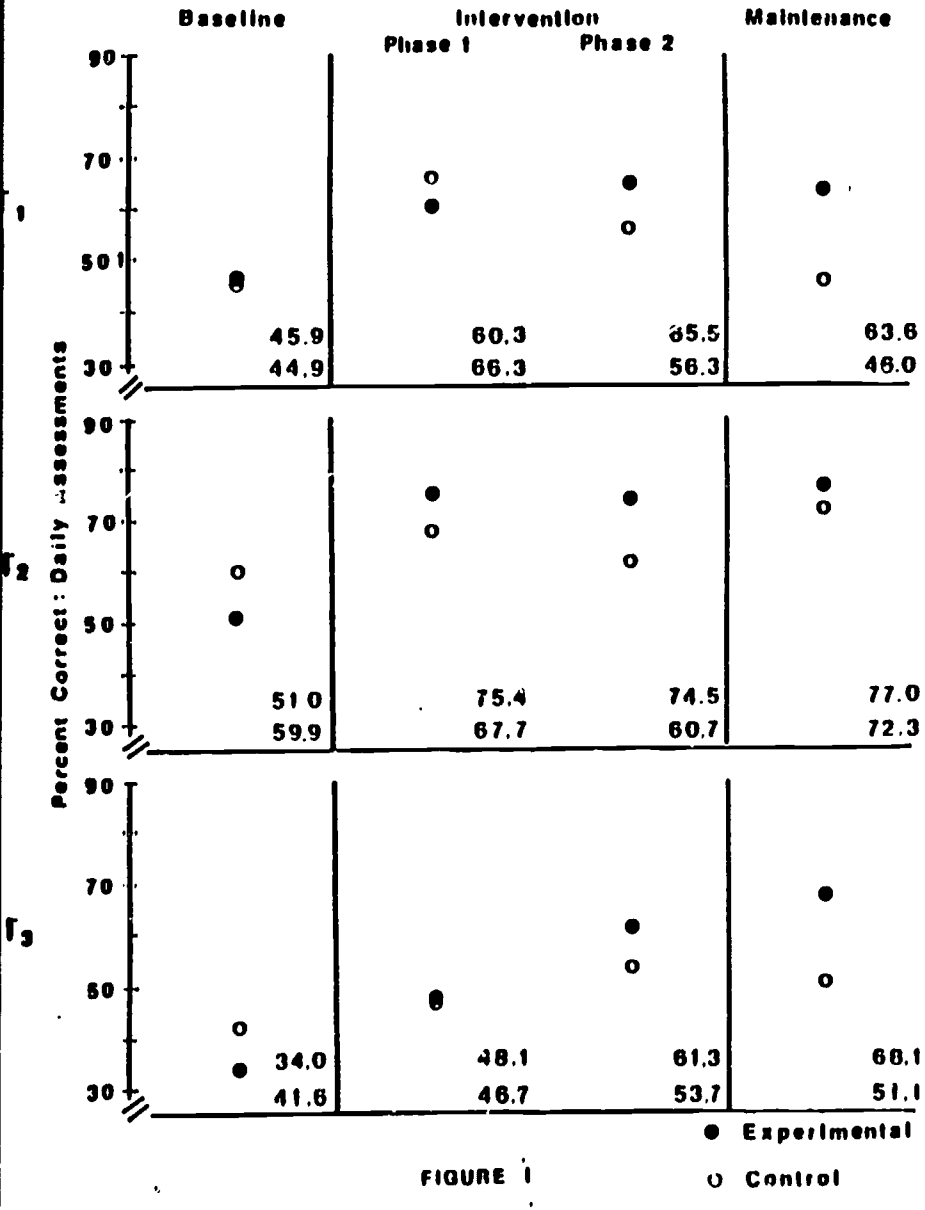
Baseline
Intervention
Phase 1
Phase 2
Maintenance
Percent Correct: Daily Assessments
90
70
50
30
45.9
44.9
60.3
66.3
65.5
56.3
63.6
46.0
51.0
59.9
75.4
67.7
74.5
60.7
77.0
72.3
34.0
41.6
48.1
46.7
61.3
53.7
68.1
51.1
Experimental
Control

FIGURE 1

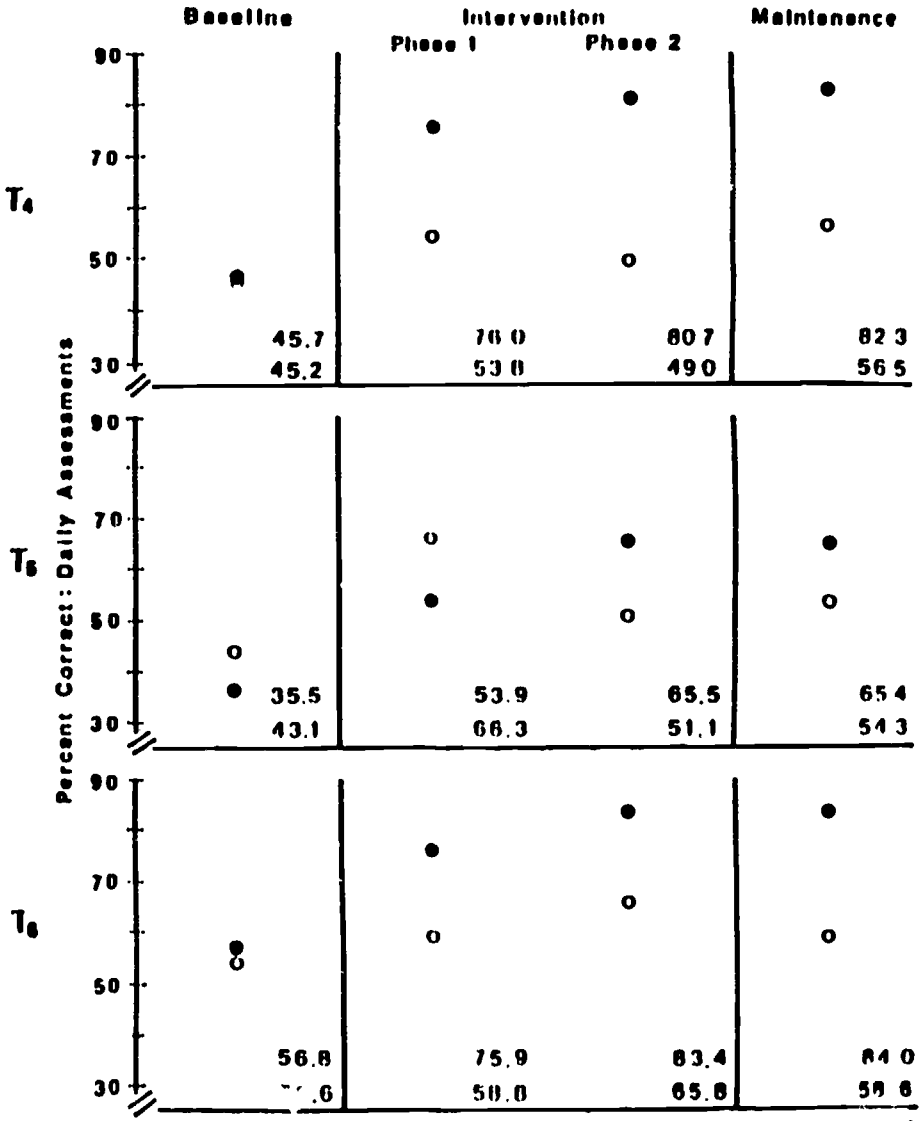
Baseline
Intervention
Phase 1
Phase 2
Maintenance
T4
T5
T6
90
70
50
30
Percent Correct: Daily Assessments
45.7
45.2
76.0
53.8
80.7
49.0
82.3
56.5
35.5
43.1
53.9
66.3
65.5
51.1
65.4
54.3
56.8
75.9
83.4
84.0
50.8
65.8
58.6
Experimental
Control

FIGURE 2