Technical Report No. 277

THE EFFECTS OF INSTRUCTION IN METACOMPREHENSION AND INFERENCING ON CHILDREN'S COMPREHENSION ABILITIES

Christine J. Gordon
University of Calgary

P. David Pearson
University of Illinois at Urbana-Champaign

June 1983

Center for the Study of Reading

TECHNICAL REPORTS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
51 Gerty Drive
Champaign, Illinois 61820

BOLT BERANEK AND NEWMAN INC.
50 Moulton Street
Cambridge, Massachusetts 02238

The National Institute of Education
U.S. Department of Education
Washington, D.C. 20208

CENTER FOR THE STUDY OF READING

Technical Report No. 277

THE EFFECTS OF INSTRUCTION IN METACOMPREHENSION AND INFERENCING ON CHILDREN'S COMPREHENSION ABILITIES

Christine J. Gordon
University of Calgary

P. David Pearson
University of Illinois at Urbana-Champaign

June 1983

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The research reported herein was supported by the National Institute of Education under Contract No. US-NIE-C-400-76-0116.

The Effects of Instruction in Metacomprehension and Inferencing on Children's Comprehension Abilities

Several large scale surveys (e.g., NAEP, 1981; Pearson & Farstrup, 1975), classrooms surveys (e.g., Guszak, 1967; Raphael, Winograd, & Pearson, 1980), and comprehension studies (e.g., Pearson, Hansen, & Gordon, 1979) have found that students obtain higher scores on literal comprehension test items (for which there is explicit textual link between question and answer) than they do on inference items (for which students must integrate information across text segments or fill in missing information).

Furthermore, surveys of questions asked by teachers in classrooms (e.g., Guszak, 1967) and available in basal reader manuals (e.g., Hansen, 1981) suggest that literal questions are more abundant in children's question-answering environments than are inference questions. Hence it is indeterminate whether the gap between literal and inferential performance is a function of an inherent task difficulty difference or a simple artifact of instructional history.

Whatever the reason, the gap exists. We would like to see it narrowed. A major purpose of the present study, therefore, was to evaluate different strategies for improving students' comprehension skills, particularly their inference ability. In order to determine what sorts of strategies might conceivably prove effective, we examined the research on basic processes in comprehension generally and inference particularly. There, we felt, might lie some clues regarding possible instructional interventions that might lead to better inference performance.

One well-developed line of research suggests that comprehension and inference are strongly influenced by the knowledge structures (schemata) that readers bring to a text. From the impetus provided by theoretical

work (Rumelhart, 1977; Anderson, 1977; Rumelhart & Ortony, 1977; Rumelhart, in press) explicating the nature and functions of schemata, several researchers (e.g., Anderson, Reynolds, Schallert, & Goetz, 1977; Anderson, Spiro, & Anderson, 1977; Steffensen, Anderson, & Joag-Dev, 1980; Pichert & Anderson, 1977; Pearson, Hansen, & Gordon, 1979; Johnston & Pearson, 1972) have demonstrated that comprehension and recall generally, and performance on inferential tasks specifically, are strongly influenced by the content schemata (knowledge about the topic addressed in the text) that students bring to the text. Corroborative evidence has been provided by other researchers not working so directly within a schema-theory framework but who are nonetheless interested in the relationship between the activation of knowledge structures or "set" and comprehension or recall (Bransford, Barclay, & Franks, 1972; Bransford & Johnson, 1973; Bransford & McCarrell, 1974) and by researchers more directly concerned with the development of inference ability across age (Warren, Nicholas, & Trabasso, 1978; Omanson, Warren, & Trabasso, 1978; Brown, 1978).

Taken as a group, these theoretical and empirical perspectives paint a picture of a very active rather than a passive reader--a reader who filters text information through a schematic filter at the point of encoding information into memory in order to produce a memorial representation in which the information presented in the text becomes largely indistinguishable from information provided by pre-existing schemata. Presumably the schemata direct at least two important functions. First, schemata are used to fill in important gaps in the inevitably less-than-fully-explicit text (e.g., character motives, instruments used to perform actions, reasons for the occurrence of an historical event). Second, schemata direct what Trabasso and his colleagues call text-connecting inferences (cf. Nicholas & Trabasso, in press); that is, sometimes the task is not so much to fill in gaps in the text but to recognize that the information presented in one proposition is related to information in another proposition by virtue of causality, sequence, enablement, attribution, etc. Furthermore, these two functions, slot-filling and text-connecting, are not limited to the point at which information is encoded into memory; they can also operate at the point of retrieval (Spiro, 1977; Pichert & Anderson, 1977).

Another, almost independent, line of theory and research implicates a different sort of schematic influence on comprehension and recall. Whereas the previous work emphasizes what might be called content schemata (knowledge about topics), an equally well-developed literature had demonstrated the influence of structural schemata (knowledge about how texts are typically organized) on comprehension, recall and inference. While there have been some attempts to characterize the broad structure of expository texts (e.g., Meyer, 1975), or text in general (Grimes, 1975; Kintsch, 1975), most of the work on text structure has focused on the narrative form.

Several formalisms, dubbed story grammars, have been developed to represent the overall structure (macrostructure) underlying a whole text. Thorndyke (1977) found that both structure and content played a crucial role in comprehension but they could function independently of one another. Passages with more familiar content were more easily understood; highly inferrable narrative structures were more readily detected and used to encode passage content. Stein and Glenn (1977) and Mandler and Johnson

(1977), although they found some developmental differences in categories of story information recalled by children, concluded that certain nodes in the story grammar were critical for story generation. Further support that the internalized story structure is used to guide comprehension and recall comes from studies that find when stories violate canonical story grammar (Stein & Nezworski, 1978), students recall protocols become reorganized according to an ideal story structure. Studies have also shown incidental learning (Bower, 1974, 1976; Thorndyke, 1977) of plot structure to transfer to the recall of similarly organized stories. These findings have led Cunningham and Foster (1978) to promote the story grammar as an instructional technique for story analysis. The intent is to get children to internalize the grammar and then use it for subsequent story comprehension and recall.

Most relevant to the present study, with its focus on comprehension and inference, are findings which suggest that story grammars allow for prediction of the types of information that will be easily recalled (e.g., high level nodes such as setting, goal and resolution), often omitted (e.g., low level actions not central to the path from goal to resolution), or easily inferred during probed recall (internal responses and character motives). In short, story grammars account for what does and does not become part of memory.

A somewhat different line of research regarding inference has been pursued by Paris and his colleagues (Paris, 1975; Paris & Upton, 1976; Paris & Lindauer, 1976). Their work begins with the same observation made by Trabasso and his colleagues, i.e., older children spontaneously draw more inferences than do younger children. However, their work suggests that this difference may be more influenced by cognitive strategies than by knowledge base. Specifically, Paris and Lindauer (1976) found that when younger children were encouraged to act out the completion of a task specified in an implied instrument sentence (e.g., He opened the door), they performed just as well as children given the explicit instrument sentence (He opened the door with a key) when later asked to name the instrument used. Younger children not directed to act out the task were less able to infer the instrument. On the other hand, for older children, the sentence input (instrument versus no instrument) did not differentiate their performance on the instrument naming task. Such data have led Paris and his colleagues to conclude that younger children are simply not predisposed to draw inferences automatically; however, when they are forced into a conscious control mode (to borrow a distinction from Posner and Snyder's (1976) work), they can do so.

Brown (1978) seems to take a middle position on the knowledge structure versus cognitive strategy issue, arguing that both help to explain the developmental increases in inference performance. However, she points out that the onset of metacomprehension abilities (i.e., the planful awareness and control of comprehension strategies) is more a function of task complexity and particular strategy than age. And she and her colleagues (Brown & Barclay, 1976; Brown & Campione, 1977) have demonstrated that children can be trained to use memory and metamemory strategies to increase their memory for new stimuli. These findings suggest that, depending upon the complexity of the task, children can be trained to use strategies that do not necessarily occur to them intuitively.

Implicit in these findings are some competing hypotheses about instructional approaches that might improve children's inference ability. The schema theory perspective suggests (for both content and structure schemata) the possibility that a treatment which heightens children's awareness of the value of invoking relevant background knowledge about the topic of selection and/or the underlying structure common to all stories (i.e., story grammars) should increase their general recall for and comprehension of stories, particularly for measures which focus on inferential behavior. On the other hand, the inference and metacognitive work more directly implicates an approach which focuses on the process of how one makes an inference and how one decides that an inference one has made is reasonable (what data sources do you use, how do you use that data to justify an answer not directly stated in the text, what are the range of allowable inferences that satisfy the constraints of the text, the questions, and both inductive and deductive logic; how big a "leap" can one make from the text?). In short, the focus is on processing (the hows) and control (what is reasonable) strategies.

Given these two perspectives, we designed an eight week instructional experiment to evaluate which of these two approaches would yield the greater growth in inference-making ability. One group, the Content and Structure group, received systematic attention, story after story, to the importance of invoking background knowledge appropriate for story understanding and to eliciting an integrated understanding of the story in terms of salient components within a story grammar (i.e., setting, problem, goal, attempts, internal responses, and resolution). A second group, the Inference Awareness group, received a treatment in which the process of drawing inferences from text and evaluating the plausibility of inferences drawn was made explicit. In addition we used a placebo control group that participated in languaging activities that, while unrelated to the inferencing, were equally as novel as the two treatments to the students in the population.

Since we were evaluating two equally plausible extrapolations from current theory and research, we had no a priori hypothesis about which of the two experimental treatments would elicit greater comprehension gains. However, we did expect that both would show greater gains than the placebo control group, especially on items tapping inferential behavior.

Method

Subjects

The subjects were 42 students who were in the "top" reading groups using the same basal reader in three fifth grade classes in one elementary school in a lower-middle class area of Calgary, Alberta, Canada. Teachers classified each child as a high or average achiever on the basis of past performance in reading (teacher assessments and standardized reading test results). Simultaneous scheduling of reading periods in all three classes enabled random assignment of the high and average readers to one of three groups. Each group was then randomly assigned to a treatment condition. Pretests administered also revealed no significant difference among the groups in performance on measures of verbal ability, standardized reading tests, written recall, experimenter-designed tests, and prior knowledge tests. Nonetheless, several of the measures were used to remove variance associated with pre-experimental "status" in subsequent regression analyses of the outcome measures.

Instructional Procedure and Materials

The experimenter met with each group for a one-half hour session per day. The first ten minutes were devoted to the differential treatment; the remaining, to a common basal reader lesson. Eight stories were taught with five instructional periods devoted to each selection. Teacher-pupil interactions (in the two experimental groups as well as the control group) in the first 10 minutes of each period were limited to 10. Each interaction consisted of one probe posed by the teacher and one student response. The responses were classed as correct or incorrect. If the response was incorrect the teacher repeated the probe and asked another student to provide a response. If the second student experienced difficulty in responding correctly, a third student was asked for a response. Analysis of taped lessons revealed that the need for a third probe did not arise frequently, nor did second and third probes occur more frequently in one treatment group than another.

Both treatment procedures and the common basal lesson were based on eight basal reader selections of the series (Ginn & Co., 360 Reading Series, 1972) normally used in the school. Five of the stories (all from one unit) were classified by the publisher as myths and folktales related to the sea. Since folktales and myths have the most similar (and fairly clear) structural characteristics compared to other types of narrative prose, they were chosen as the first five selections. The remaining three were "modern narratives" all with a theme related to a specific country. The last three selections were more varied in structure. Follow-up assignments for all groups were correlated with daily lessons common to all and based on accompanying series workbooks and basal materials.

One worksheet per week, while containing a core of common questions, was specifically tailored to allow additional practice consistent with each treatment.

Independent Variables

The main independent variable was the type of differential treatment each group received daily for 10 minutes. Prior to reading the story, the Content and Structure group received, on the first teaching session related to a story, one of a variety of activities to expand their existing content schemata on the topic.

Selection of content for background development was based on a set of criteria and presented as a block of information in the ten minutes of the first teaching session. The background information to be developed and/or expanded was selected on the basis of the following criteria: (a) relationship to the theme as depicted in the teacher's guide, (b) the propositions high in the story structure, and (c) world knowledge about actions, objects, and events specified in text (Warren, Nicholas, Trabasso, 1978). An attempt was made to include in the block presentation some of the same information that would be generated and used by the inference-awareness group within the context of reading the story during the course of their experimental lesson. This aspect of instruction, however, was not systematically controlled. It did, however, provide a core of information on which to base judgments about performance the common scriptally implicit questions (cf. p. 13) for the weekly worksheets. Table 1 lists the range of activities included in building content schemata.

Insert Table 1 about here.

In the four 10-minute sessions that followed, the Content and Structure group received instruction aimed at developing text-structure awareness. The story was read, section by section (i.e., a minor setting, and an episode), to determine the macrostructure of the narrative and note it down on a schematic representation (see Figure 1). The appropriate superordinate statements for each category were elicited orally from the group. Children "filled in" each slot of the diagram by writing in the determined proposition.

Insert Figure 1 about here.

Typical probes consisted of story structure related questions such as:

1. What is the major setting of the whole story?
2. What is the main goal or problem in the story?
3. What is the setting of the first episode?
4. What started the chain of events in the first/second, etc. . . . episode?
5. What does the character think, feel or want?
6. What does the main character do?
7. What are the reactions to this action?
8. What happens as a result of the main character's action?

Typically, children's responses (which were paraphrases of text) for setting, (1) above, for example, would consist of statements such as:

> Ramon and his family were Mexicans who made their living selling pottery.

In an effort to limit teacher-pupil interactions to 10 per day while at the same time completing the whole story structure in four lessons, students were provided with partially completed schematic representations (i.e., in a "macrocloze" fashion some categories were completed with information from the story, others were left blank for completion during the classroom interactions).

Prior to attempting to determine the story schema of the long narratives in their reader, one instructional session was devoted to familiarizing the children with story schema and the schematic representation by using a simple narrative (Guthrie, 1977). No emphasis was placed during the first two weeks on determining the relations between episodes [CAUSE, AND THEN]. In the experimenter's judgment, a sufficient amount of time first had to be devoted to familiarizing children with the categories within the story structure. For the remainder of the study, on the average, one question was included during the instructional session to determine relationships between episodes. The intent was to help children to perceive that episodes within the story plot are linked.

No attempt was made to make the stories correspond to the Stein and Glenn (1977) model as not all postulated categories within an episode were present in the stories analyzed (i.e., initiating events, reactions resolutions). Similarly, other episodes digressed from the ideal story schema and contained several reactions or an action-reaction-action-reaction sequence within the episodic structure. The eight stories differed in total number of episodes. Nevertheless a basic common structure was apparent in the majority of episodes within stories and over all selections taught.

Prior to the initiation of the study, the macrostructure of each narrative had been determined according to the following criteria:

1. The experimenter developed superordinate statements that represented information in the higher order nodes of a story schema. (The stories were far too lengthy to be parsed according to individual propositions.)

2. The superordinate statements for the eight stories were classified into the appropriate category specified by Stein and Glenn's (1977) story grammar. Interepisodic relations were specified. Adaptions in format of the grammar (for school use) were based on Cunningham and Foster (1978). Further, the categories within an episode were titled as follows: initiating event = starter; internal response = goal, thought, feeling or plan; attempt = action; consequence = what happens; reaction = reaction.

3. The analyses were presented to another rater who disagreed or agreed with the parsing.

4. When a disagreement occurred between the two raters, the narrative and the proposed parsings were presented to a third rater familiar with narrative macrostructure analysis.

5. The particular superordinate statement was assigned to the category for which two out of three raters agreed.

6. If the third rater was not in agreement with any one of the other two raters on a particular story segment, he presented his own categorization. On the average, the third rater disagreed on no more than 1 or 2 statements per story.

7. Having the input of the third rater, final agreement on the parsing was reached by the first two raters by mutual consent.

Initial inter-judge agreement of 91% was found. The second reliability check yielded an inter-rater agreement (among the three raters) of 95%.

The instructional treatment received by the Inference-Awareness Group consisted of making children aware, through a step-by-step procedure, of a strategy that they could employ consciously to make inferences. This awareness was developed in the context of reading the relevant passages. The step-by-step modeling and feedback techniques (see Table 2) was developed by utilizing ideas gleaned from Davis (1978) and Pearson and Johnson (1978).

Insert Table 2 about here.

An example of a partial lesson to teach the inference-awareness strategy is presented. The following teacher-pupil interaction relates to generating the relevant schemata and textual elements in order to make a script-based inference. It is based on the procedure for week 5 (Table 2).

1. The experimenter read the relevant part from the text.

   "Sons of the Northwind" she addressed them, "halt your flight. The harpies cannot be killed by sword. Because you have been brave enough to pursue them, the prophecy will come true. I swear to you by the River Styx--and no God can go back on that oath--I swear that the Harpies will never bother Phineus again. Return to the feasting." So she spoke and flew back to high Olympus, the home of the gods. (Clymer, Gates, & McCulough, 1972, p. 254)

2. The experimenter stated the relevant clues:

   a. From the story--the Messenger said, "I swear to you by the River Styx."

   b. From the head--Styx is the river that separates the land of the living and the dead.

      --The Bible is considered sacred.

3. The experimenter asked for the inference as to why swearing on the Bible would be the same as swearing the way the messenger of Zeus did.

The next teacher-pupil interaction focuses only on relating textual elements (from week 5). Using the same text as in (1) above but with a focus on text-connecting inferences, the interaction in week 5 (Table 2) would proceed as follows:

4. The experimenter stated the relevant clues:

   a. From the story--The messenger said, "Halt your flight" and "Harpies cannot be killed by sword."

   b. From the head--none.

5. The experimenter asked for the inference as to why the messenger told the Sons of the Northwind to "Halt their flight?"

Although an attempt was made to equalize the number of text-based and script-based inferences made daily, the type of information presented in the story sequence sometimes dictated the balance. The same criteria as that utilized in the Content and Structure Group was used as a basis for determining which information from prior knowledge should be generated in the context of reading the selection and thus applying the strategy.

Proficiency of students at steps a), b), and c) of Table 2 was determined on the basis of the weekly worksheets which, with the exception of the core of common questions, were specifically related to each group's differential treatment. When 90% of the students correctly completed six out of seven questions pertaining to their treatment, all students proceeded to the next step on Table 2. Such a procedure dictated that no a priori assumptions could be made on the amount of time that would eventually be allotted to steps d) and e) on Table 2.

The treatment received by the Control Group consisted of a variety of literature, language or creative activities related to the basal selection. The major rationale behind this treatment was to ensure that the Control Group received a treatment as novel as and as interactive as that of the other two groups.

Dependent Variables

Comprehension questions following instructional selections. One type of dependent measure, the instructional story worksheet, was administered at the completion of each selection. The first set of three questions (one each of textually explicit, textually implicit, and scriptally implicit (based on the Pearson and Johnson, 1978 taxonomy)) were common to all treatments and had not been previously discussed in the instructional sessions. The remaining seven questions were specifically related to each group's differential treatment. All worksheets were completed under the experimenter's supervision. Access to text was permitted but no experimenter assistance on questions was given and no collaboration with fellow students was allowed. As each student completed the worksheet he or she submitted the work face-down to a designated student to return to the experimenter. On the following day of instruction, a few minutes were taken to return the graded worksheets and to review in each treatment group the section pertaining to their specific treatment. However, no discussion of the first three questions (common to all three groups) was undertaken. Correct answers were simply provided orally. The worksheets were then immediately collected again.

Responses to the three common questions (the only ones used in data analysis) were given as optimal value score (1) or a minimal value score (0) according to the following set of criteria.

1. Textually explicit answers were given an optimal value score if they were reproductions or synonyms of the target statements in the narrative.

2. Textually implicit answers were given an optimal value score if there was an appropriate match between the question statement and the response statement both of which were derived from the text. Such responses required the student to make at least one step of logical or pragmatic inferring in order to get from the question to the response. Reproductions

of and/or ideas synonymous to those found in the response statement from the text were both accepted as optimal value scores.

3. Scriptally implicit answers were given an optimal value score of one if they were based on information in the narrative (either inferrible or explicitly stated) but required information from the student's prior knowledge (scriptal knowledge) to answer the question. The appropriate answer, while based on information in text, came from prior knowledge. In other words, in order to "come up" with the answer the reader must first have read and remembered the text, and secondly brought to bear pertinent information from script. The answer therefore showed an integration of text and script.

The inter-judge agreement was 90% for scoring responses.

Written recall. Further, all children were asked to provide written recall of the last instructional selection (number 8) in the form of a story summary. Students were instructed as follows:

> Please write a summary of the story, ______________.
> That is, write the story from beginning to end in your own words. Include what you think is important. Please do your best.

The students' protocols were then compared to the validated macrostructure of the story in order to assess the presence or absence of information classed as high in story structure and to classify the information in terms of the major categories specified by the grammar. Scoring was done according to the following criteria. The propositions were scored as correct if:

1. they were accurate in terms of the semantic content established when the story grammar was developed,
2. if they were sufficiently complete to capture the gist of the superordinate statement in the story grammar in synonymous terminology.

No credit was given for:

1. partial information (i.e., deleting an important constituent (one idea out of two) in the superordinate statement),
3. information low in the story structure (i.e., details),
4. statements demonstrating confusion or a lack of comprehension.

Credit was given for complete and accurate statements which captured the essence of a category but were not in exactly the correct story sequence. Students were also not penalized for the following types of errors:

1. incorrect spelling, and
2. omission of a word(s) which did not affect the presentation of ideas in the proposition.

The inter-scorer agreement was 92%. Since some propositions in the student protocols were difficult to score and/or categorize, where differences existed, agreement on the final assignment was reached by mutual consent.

Experimenter-designed silent reading comprehension tests. Silent reading comprehension tests devised by the experimenters were administered after the completion of the study. Narrative selections, i.e., transfer stories, unfamiliar to the subjects, were selected from different basal

reader series. Stories were shortened and/or modified to equalize story lengths and to obtain the desired number of specific question-answer relations. Dale-Chall readability levels revealed that the narratives were approximately equal in difficulty level. Stories were prepared for comprehension tests given immediately following the study's completion; one each to be administered under text access (students could read the text while they answered) and no text access conditions. The delayed test, administered 16 days after completion of the teaching phase, consisted of two different unfamiliar selections, again one to be completed with reference to text available and one without such reference. Under text access conditions, the following instructions were read orally by the experimenter while students followed along silent reading the printed directions.

> Read the story carefully. You will have 15 minutes to read the story. After the reading, you will be asked to write down the answers to 15 questions. You will have 15 minutes to answer the questions. When you are answering the questions you can refer back to the story booklet on your desk.

Under no text access conditions, the following directions were provided orally while students followed along silent reading directions.

> Read the story carefully. You will have 15 minutes to read the story. After the reading, the story booklet will be picked up. You will then be asked to write down your answers to 15 questions. You will have 15 minutes to answer the questions.

After they read each selection, students completed 15 questions, 5 textually explicit, 5 textually implicit, and 5 scriptally implicit. These questions were based on the Pearson and Johnson (1978) criteria as well as the following additional considerations.

(1) Questions were related to the theme as depicted in the basal reader guides accompanying the selections;

(2) Questions tapped information related to the higher level nodes in each story structure;

(3) The types of inferences assessed by the textually implicit and scriptally implicit questions were selected on the basis of an inference taxonomy (Warren, Nicholas, & Trabasso, 1978).

(4) Test questions were keyed to types of inferences (based on the Warren et al., taxonomy) developed in the Inference-Awareness group during instructional sessions. (No systematic analysis of this aspect was made, however);

(5) Pilot testing was conducted to eliminate questions with abnormally high ($p > .85$) or low ($p < .35$) difficulty levels.

The inference taxonomy (Warren et al., 1978) accounts for some of the questions asked on the experimenter-designed test. Pilot testing resulted in certain types of inference questions being discarded in some selections but not in others. The resulting imbalance in types of questions asked (according to the Warren et al. taxonomy), therefore precluded any statistical analysis on the basis of the subclassifications in their taxonomy.

The order of probes, while reflecting the sequence of events in each story, contained a well-balanced dispersion of question types. The answers were scored according to the same set of criteria as the instructional story worksheets discussed earlier. For example, the following responses from the Renegade Husky (Durr, Windley, & Yates, 1971, 1974) would receive an optimal value score of one (1).

Text:	(1)	"Nanook? Oh, I sold him to the sourdough at Balsalm Ridge. He (Jake) asked to leave the dog here for a couple of weeks to get used to the whites."
Textually explicit question:	(1)	For how long did Jake leave Nanook at the trading post to get used to the whites?
Answer:	(1)	He left him for a couple of weeks. He left him for two weeks.
Text:	(3)	He started before dawn next morning, snowshoes strapped to his back. It was eight miles to Jake's place.
Textually implicit question:	(3)	What means of transportation did Rick use to get to Jake's place?
Pragmatic Answer:	(3)	He walked. He went on foot.
Text:	(4)	As he (Rick) stood undecided he heard a low howl from behind the cabin. His heart leaped to his throat. But a howl from that breed of dog did not surprise Rick at all.
Scriptally implicit question:	(4)	Why did a howl from that breed of dog not surprise Rick?
Answer:	(4)	Nanook was a Malemute which is part wolf. He was part wolf and wolves howl. He had wilf ancestors (wolves). (The information that Nanook was a Malemute husky had been explicitly stated earlier in text and now had to be remembered by the child in order to answer the question. That Malemutes are part wolf comes from prior knowledge.)

A reliability check of this response categorization scheme yielded an inter-scorer agreement of 95%.

Subtests of the Metropolitan Achievement Test. The same form of the Word Knowledge (vocabulary) and Reading (comprehension) subtests of the Metropolitan Achievement Test, Intermediate level, Form F, 1971 edition, as administered for the pretest, were administered as a posttest. The subtests were administered and scored according to the directions of the test publisher.

Pretests

In order to achieve more experimental precision in the analyses of the dependent measures, data was collected prior to the experiment on the following tasks: The Metropolitan Achievement Test, a free recall of a story, prior knowledge tests concerning the topics addressed in the post-test selections, and an answering questions test under both access and no access conditions) that was identical in structure to the immediate and delayed post-test task.

Data Analysis Procedures

There were four sets of post-test data available for analysis: (a) comprehension as measured by instructional worksheets, (b) comprehension as measured by the four (2 immediate and 2 delayed) transfer stories, (c) scores on the Metropolitan Achievement Test, and (d) recall on the last instructional story.

The general plan for our hierarchical regression analyses was as follows. First, we removed variance associated with any appropriate measures collected as pretest data (e.g., score on Metropolitan Achievement Test, pretest recalls, etc.). Second, we partitioned the variance

associated with treatment into two contrasts. For all but the recall data that meant contrasting the two experimental groups with the control group (usually Contrast 1) and then comparing the two experimental groups (usually Contrast 2). For the recall data, Contrast 1 compared the Content and Structure group with the other two, and Contrast 2 compared Inference Awareness with the Control. Our rationale for this deviation from the general plan was that only the Content and Structure group received instruction that should influence recall behavior; in this instance, therefore, the Inference Awareness group was more like a control group. Third, we examined interactions between the contrasts and the covariates that had been removed in Step 1. Fourth, where appropriate, we conducted a within subjects regression to examine the effects of question type, especially in interaction with treatment.

For each of the four sets of data, a preliminary regression analysis was conducted to (a) determine whether or not treatment interacted with any of the appropriate pre-test covariates (e.g., pretest Metropolitan, pre-test answers to story questions, recall on the first instructional story) and (b) where appropriate, to determine whether or not question type interacted with treatment and/or any appropriate covariates.

In three cases (transfer stories, last story recall, and post-test MAT combined scores) there were significant covariate by treatment interactions; hence, separate regressions were computed for each treatment, using the appropriate covariate to predict post-test performance for each treatment group. In each case, we conducted Johnson-Neyman (Kerlinger & Pedhazur, 1980; p. 256-258) tests to determine at what points in the distribution of covariate scores the treatments differed from one another on the dependent measure. In only one case (instructional worksheet questions) did question type interact with treatment. Also in that instance, treatment did not interact significantly with any covariates. Hence, we were able to conduct a treatment by question type analysis of co-variance, adjusting for pre-test scores on the Metropolitan Achievement Test.

Once these overall analyses were completed, we proceeded to take a finer look at the data to see if we could pinpoint the sources of our main and/or interaction effects; specifically, we looked at each of the four transfer stories separately and we examined last story recall by various recall categories.

Any time that question type was a factor in our statistical design, it proved to be a significant source of variance. However, we opted not to analyze the effect of question type separately from its interaction with treatment and/or covariates. We took the position that there was no way of controlling the inherent difficulty of questions; in other words, it was difficult, if not impossible, to write questions sets such that a correct response was equally probable for questions classified as text-explicit, text-implicit, or script implicit. Hence, any interpretation of differences attributable to question type could be misleading. This situation, however, does not rule out the possibility of examining the interactions of question type with other factors since all treatment groups received the same questions on the post-test measures. Therefore, we have chosen to report, but not interpret, question type main effects.

For all analyses, we have reported cell and marginal means and standard deviations. In addition, for the sake of those readers more used to analyses of mean scores, we have reported the means when they were

adjusted for those covariates removed in Step 1 of the analysis. While these adjusted means never, in any real sense, entered into the analysis, they are provided to give the reader a sense of what was being compared when treatment contrasts were analyzed. They are there only for convenience; no interpretation should be placed on them save in the one instance (instructional worksheets) in which we were able to conduct an analysis of covariance.

Results

Worksheets Following Instructional Selections

For this analysis, the three questions common to all groups on the instructional worksheets following each story were pooled, by question type, to create a set of eight items per question type. The between subjects regression effects are summarized in Table 3. Note that the only effect accounting for a significant source of variation was the MAT test administered prior to the experiment. It accounted for 32% of the overall variance, $F(1,40) = 19.09$, $p < .001$. Marginally significant effects were found for pretest scores on the experimenter designed test, $R^2 = .06$, $.05 < p < .08$, Contrast 1 (Content and Structure + Inference Awareness - Control), $R^2 = .05$, $.05 < p < .07$, and the interaction of MAT score with Contrast 2 (Contrast 2 compared Content and Structure against Inference Awareness), $R^2 = .05$, $.05 < p < .07$. Following our plan, we conducted an analysis of covariance using treatment as a between-subjects variable, question-type as a within-subjects variable, and MAT score (the best predictor) as a covariate. This analysis yielded a marginal effect for treatment, $F(2,38) = 2.933$, $.05 < p < .07$, a significant question type

---

Insert Table 3 about here.

---

effect, $F(2,77) = 55.51$, $p < .001$, and a significant interaction between treatment and question type, $F(4,77) = 3.47$, $p < .02$.

An inspection of the adjusted means (see Table 4 for both unadjusted and adjusted means) reveals the probable source of the interaction: Patterns of responses on both text-explicit and text-implicit questions reveal few reliable differences among the treatments. However, on script-implicit questions, there appears to be a definite Inference Awareness > Content and Structure > Control ordering on the adjusted means. In fact, the simple effects tests for just the script implicit questions corroborate such an ordering of adjusted means ($p < .05$).

Insert Table 4 about here.

Recall that the Inference Awareness group received direct explicit training on how to answer precisely these sorts of questions. At least within the context of answering new inference questions from the same stories within which instruction was embedded, students who received inference awareness training appeared to benefit on what proved to be a difficult set of questions (note the overall unadjusted mean of 3.07 out of 8 on this measure).

Questions from the 4 Transfer Stories

In the first analysis of these data, we conducted an overall regression analysis on answers to all four stories, looking for both between subject and within subject sources of variance worthy of further exploration. The between subjects portion of the analysis included main effects tests for two covariates, MAT total score and score on the pretest stories, two contrasts comparing, respectively, the two experimental groups with the control and one experimental group with the other, and the interactions of

the contrasts with the covariates. The within-subjects analysis (using dummy coding for question type and story) examined main effects for question type and story, the interactions of question type and story, and the interactions of these within-subject variables with the two between subject contrasts.

Overall regression. In the between subject analysis (see Tables 5 & 6), we found significant main effects for MAT, Pretest, and Contrast 1. However, we also found that Contrast 1 interacted significantly with pretest score, suggesting that we analyze the effects of pretest separately for each level of Contrast 1.

Insert Tables 5 and 6 about here.

The within subjects variance (total explained variance = .75) revealed that both question type ($R^2 = .44$) and story ($R^2 = 13.3$) accounted for significant sources of variance as did the interaction of the two ($R^2 = 10.11$). The sum of the R^2 for all the interaction tests between various combinations of these within subjects variables with Contrasts 1 and 2 was .0674, indicating little hope of explaining results with these higher order interactions.

Given this state of affairs, we decided that the most sensible analysis plan involved computing separate regressions for the two levels of Contrast 1 (the two experimental groups and the control group), using pretest score to predict post-test score. Then to check the stability of the results, we would conduct the between subjects analysis for each story separately, adding the prior knowledge test for each story as a covariate. We decided that the small amount of variance associated with the interactions of the contrasts and question type did not merit further examination of question type as a factor.

Separate regressions. The separate regression equations for the two experimental groups (Inference Awareness and Content and Structure) and the control group appear at the bottom of Table 5 and are displayed graphically in Figure 2. They suggest a differential effect of treatment as a function of pretest score. One substantive interpretation is that as pretest score increases, either of the two experimental treatments has an increasingly beneficial effect on post-test performance. In short, we may have discovered two treatments that uniquely benefit higher achieving students. The application of the Johnson-Neyman technique supports such an interpretation; for any pretest score above 16, the difference between experimental and control students is significant (there is no point, by the way, above a pretest score of 0, at which control students outperform experimental students significantly). There were 13 in the 2 experimental groups and 5 in the control group who achieved pretest scores above 16.

While this substantive interpretation seems reasonable in light of the shape of the regression lines and the application of the Johnson-Neyman technique, there is an artifactual interpretation that can be raised. It is possible that there was a floor effect on the pretest, the post-test, or both, and that scores are so unreliable at the lower end of each measure that we were unable to detect any real differences among the treatments. Hence, scores for low-scoring students in the control group are unreliably inflated on the post-test causing the appearance of an aptitude by treatment interaction that does not really exist. We mention this interpretation only

----
Insert Figure 2 about here.
----

in the interest of conversatism; we are comfortable with the substantive interpretation of an aptitude by treatment interaction suggesting unusual benefit for either experimental group beyond a pretest score of 16.

Regressions for each of the four stories. The regression effects for each of the four stories examined separately appear in Table 7. What is interesting here is the differential effects of Contrast 1 (the two experimental versus the control) and Contrast 2 (Content and Structure versus Inference Awareness) as a function of text access while answering questions. Notice that C_1 is a significant source of variance for the text access story given immediately after the experiment ($R^2 = .08$), and approaches significance on the delayed text access story ($R^2 = .04$). However, Contrast 1 disappears when access to the text is removed, while Contrast 2 approaches conventional levels of significance in these two cases ($R^2 = .03$ and .02, respectively). There are two interactions of note in these four analyses. The Pretest by Contrast 1 interaction shows up on the delayed text access story and the prior knowledge test interacts with Contrast 2 on the immediate no text access story. We computed (but have not graphed) the appropriate separate regressions for these interaction effects.

Insert Table 7 about here.

As one might predict, the regressions on the delayed text access story match those found on the overall analysis (Figure 2), perhaps pinpointing the source of the interaction effect found in the overall analysis. On the immediate no text access story, the separate regressons for each of the two experimental groups yield little interpretable data of interest.

Total Recall on Last Instructional Story

The regression analysis here involved using MAT total score and total recall on the first instructional story as covariates. Contrast 1 compared the Content and Structure with the other two; Contrast 2 compared the Inference Awareness group with the Control. After the overall regression was computed, we examined each category of recall in order to pinpoint effects.

For the total recall analysis (see Table 8), MAT score ($\underline{R}^2 = .231$) and total first story recall ($\underline{R}^2 = .129$) both explained significant amounts of variance, as did C_1 ($\underline{R}^2 = .184$) and the interaction of first story recall with C_1 ($\underline{R}^2 = .047$). Separate equations were computed for each level of Contrast 1, using total first story recall to predict total last story recall (the equations are reported in Table 8 and graphed in Figure 3). The application of the Johnson-Neyman technique to these separate regressions revealed that last story scores for the Content and Structure group exceeded those for the Inference Awareness plus Control group any time the pretest score exceeded 19 ($\underline{N} = 14$); conversely, scores for the Inference Awareness plus Control group exceeded those of the Content and Structure group whenever the first story score was less than 4 ($\underline{N} = 6$). So in this analysis there are two zones in which significant treatment differences emerge. For the few students with very low first story scores, the Content and Structure treatment elicited significantly poorer last story recall; for moderately high or higher scores, significantly greater recall.

Insert Table 8 and Figure 3 about here.

When regression analyses were performed separately for each of the eight categories of recall (see a summary of significant effects only in Table 9), the following pattern emerged. First, on two categories (major setting and internal response), no effects explained significant amounts of variance. Both status variables (MAT and first story recall) explained significant amounts of variance on attempts (actions designed to lead to goal fulfillment) and consequences (outcomes of attempts). In addition, initial story recall entered into the equation for minor setting recall, and MAT entered that equation in the form of an interaction with C_1. Contrast 1 (C_1) proved to be a significant source of variation on all six of the dependent measures in which significant amounts of variance were explained (total explained variance on these six measures ranged from .304 to .599); and, on two measures, it was the only significant source of variation. The $\underline{R}^2$ for C_1 ranged from a low of .078 for attempts to a high of .225 for minor setting. Contrast 2 (C_2) never proved significant in any of these analyses. Thus, it seems reasonable to conclude that the training received by the students in the Content and Structure Schemata group facilitated total story recall by increasing their general ability to store and retrieve the various components that comprise what we think of as "story structure." The overall interaction of First Story Recall with C_1 compromises this interpretation to a certain degree (it limits the interpretation to those students with first story recall scores above 19); however the fine grained analysis did not shed any light on the specific source of this overall interaction since the First Story Recall by C_1 interaction failed to explain a significant amount of variance on

---

Insert Table 9 about here.

---

any of the specific measures. The consistent pattern of effects for C_1 leads to the conclusion that the students learned what we taught them. Further, in light of the failure of C_2 (i.e., the Inference Awareness group never exceeded the Control) to emerge in any of these analyses, the data suggest that simply forcing students to relate new story information to known prior knowledge via inference awareness training does not facilitate story recall to nearly the same degree as does helping students to relate new to known information (the content part of the Content and Structure training) within the context of a structural framework for understanding and remembering stories (the structure part of that training).

Metropolitan Achievement Test Scores

Using the same test given as a pretest as the covariate, a regression analysis was conducted on the posttest scores for the sum of the vocabulary and comprehension subtests of the Metropolitan Achievement Test. The regression analysis (see Table 10) indicated only two significant sources of variation: MAT pretest (R^2 = .73) and the interaction of Contrast 1 (Inference Awareness and Content and Structure versus Control) with the MAT pretest score (R^2 = .03). Separate regressions were computed (see Table 10), graphed (see Figure 4), and subjected to the Johnson-Neyman technique. For pretest scores below 48 (N = 7), post-test scores for Control subjects are superior to those for the two experimental groups; for pretest scores above 63 (N = 16), scores for the experimental groups are superior to those for the Control group. As was the case with total recall, a cross-over interaction is indicated; either of the two treatments

---

Insert Table 10 and Figure 4 about here.

---

was beneficial for high ability students whereas the control treatment was beneficial for low ability students.

Discussion

The Aptitude by Treatment Interaction Issue

Taken together, these data suggest that both experimental treatments were especially beneficial to students who, on one pre-experimental measure or another, exhibited higher aptitude for reading. The net result of our eight weeks of training was to show increasing gains as a direct function of how well students were already performing. Conversely, for students who performed relatively poorly on comprehension tasks prior to the experiment, the control treatment was either more or equally beneficial.

Were it not for the consistency of these aptitudes by treatment interactions, one could raise the artifactual explanation of unreliable scores at the low end of the performance continuum concealing a floor effect that our measures were unable to detect. The consistency remains, however, even on a measure like the Metropolitan Achievement Test, which is less vulnerable to charges of unreliability than are our experimenter-designed measures. So, we are left with the aptitude by treatment interaction argument. Either of the experimental treatments, Inference Awareness or Content and Structure training appears to be helpful to students in improving their general comprehension abilities, as measured by MAT score or by the sum of text-explicit, text-implicit and script-implicit questions on transfer stories, provided that students were already achieving fairly well on some general measure of comprehension.

When comprehension is measured by story recall, the aptitude by treatment interaction explanation arises again. Only this time it is more treatment specific. Only the high-aptitude Content and Structure students appear to benefit from the instruction provided; students in the Inference Awareness group are indistinguishable from Control students on this measure.

The question posed by these data is why the experimental treatments should be more helpful to higher than lower ability students. Intuition (high ability students discover these strategies on their own and low ability students need the help more) as well as evidence gathered by Hansen and Pearson (in press) on a technique not altogether dissimilar from our Content and Structure training argue for exactly the opposite kind of aptitude by treatment interaction. Further our training did include features such as pupil responding, correction procedures during instruction, feedback on worksheets, and proof of meeting criteria of proficiency before proceeding to the next step. Judging from teacher effectiveness literature (see Duffy, 1980) such procedures tend to increase the chances of obtaining intervention gains with lower level ability students. Our treatments, therefore, should have proven uniquely beneficial to the lower ability students. Since they did not, two possible arguments arise.

It may be that either type of experimental training is sufficiently complex that some minimal threshold of achievement must be reached before it can prove helpful to a student. Given what we discovered in trying to implement the training, this argument seems more plausible for the Inference Awareness group than for the Content and Structure group. It was not always easy for students to see the relationship between text evidence and plausible inference answers. But there was nothing very mysterious

or complex about introducing and discussing key concepts or trying to place important story elements into a relatively simple story framework.

The alternative argument hinges on the difficulty of the training materials. Perhaps the stories that the students were reading were relatively more difficult for the lower ability students than for the higher. In a situation where the lower ability students were grappling both with difficult content and novel instructional devices, the cognitive load was simply too much for them to handle.

Our data and experimental design are incapable of distinguishing between these explanations; the answer awaits further research.

<u>Specificity of Transfer</u>

We had expected to find that our experimental treatments would have their greatest effect on inference questions, either text-implicit or, most predictably, script-implicit questions. The within-subjects regression analyses on the four transfer stories taken together or analyzed separately show precious little evidence of treatment by question type interaction. The only indication of such an effect is the question type by treatment interaction from the analysis of covariance on the questions from the instructional worksheets. Recall that on that measure the Inference Awareness group showed a peculiar advantage on script-implicit questions. Since these questions were taken from the stories in which the instruction was embedded, one can conclude that students in this group were able to use their training in the short-term experience of the study. However, this short-term advantage soon became amalgamated with their overall approach to answering questions, at least as measured by comprehension on the four transfer stories.

The recall data present a more compelling picture of specificity of transfer. Here the group that received training on how stories are put together were apparently able to use that knowledge to put a story together. The strong main effect for the training, qualified as it must be by the aptitude by treatment interaction, suggests very specific transfer of training.

Also, we can marshall the consistent tendency (albeit statistically non-significant) for Contrast 2 (Inference Awareness versus Content and Structure) to favor the Inference Awareness group over the Content and Structure group on question-answering measures as further support for specificity of transfer in this study.

Text Access and Treatment

Recall that the effect of Contrast 1 (Inference Awareness plus Content and Structure versus Control) found for the data summed across the four stories was an important source of variance only for the two stories in which access to text was allowed during the question answering task. At first glance this might seem counter intuitive; after all, if the training is designed to promote integration of text with prior knowledge, should it not emerge more strongly in a context in which students are not able to reexamine the text? We think not. An emphasis on integrating text information with prior knowledge need not be biased toward prior knowledge. When one is searching a text for clues that support an inference or for sequencing causally linked events, one is using the text. Hence one interpretation of our Contrast 1 in text access situations only is that we taught students how to use a text more efficiently to help generate a variety of answers to a variety of questions.

Conclusions and Recommendations

We would have preferred to find that the two experimental treatments we designed, each representing an alternative view about how and why inferential comprehension can be improved, were universally effective for students of all ability levels. Then the job of making recommendations for instruction would be easy. Our data better support the conclusion that either of these approaches will improve general comprehension for higher ability students and that, as a bonus, the Content and Structure training will also help those same students recall stories better. The only data that support an across-the-board recommendation come from the instructional worksheets: If you want students to become better at drawing inferences to prior knowledge (i.e., script-implicit tasks) then train them to do so; it helps at least in the context in which you embed the training.

Otherwise we are left with recommendation that for students who are already on the right track toward good comprehension you can nudge them even further ahead with either of these techniques. We wish that we could recommend something more substantial for those students most in need of our help. At this point, we cannot. One telling question posed by this research is, How can we get students to the point where they can benefit from instruction we know works for those already beyond that point? It remains for further research to ferret out the mysteries of that subtle but important relationship between ability and instruction.

References

Anderson, R. C. The notion of schemata and the educational enterprise: General discussion of the conference. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. Frameworks for comprehending discourse. American Educational Research Journal, 1977, 14, 367-381.

Anderson, R. C., Spiro, R. J., & Anderson, M. C. Schemata as scaffolding for the representation of information in connected discourse. American Educational Research Journal, 1978, 15, 433-440.

Bower, G. H. Selective facilitation and interference in retention of prose. Journal of Educational Psychology, 1974, 66, 1-8.

Bower, G. H. Experiments on story understanding and recall. Quarterly Journal of Experimental Psychology, 1976, 28, 511-534.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretative approach. Cognitive Psychology, 1972, 3, 193-209.

Bransford, J. D., & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Bransford, J. D., & McCarrell, N. S. A sketch of a cognitive approach to comprehension. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes. Hillsdale, N.J.: Erlbaum, 1974.

Brown, A. L. Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology. Hillsdale, N.J.: Erlbaum, 1978.

Brown, A. L., & Barclay, C. R. The effects of training specific mnemonics on the metamnemonic efficiency of retarded children. Child Development, 1976, 47, 71-80.

Brown, A. L., & Campione, J. C. Memory strategies in learning: Training children to study strategically (Tech. Rep. No. 22). Urbana: University of Illinois, Center for the Study of Reading, March 1977.

Clymer, T., Gates, D., & McCullough, C. The sun that warms. Toronto: Ginn & Co., 1972.

Cunningham, J. W., & Foster, E. O. The ivory tower connection: A case study. The Reading Teacher, 1978, 31, 365-369.

Davis, J. E. Developing inferential comprehension. Session presented at the International Reading Association, Houston, Texas, 1978.

Duffy, G. G. Teacher effectiveness research: Implications for the reading profession. Directions in Reading: Research and Instruction, 30th Yearbook of the National Reading Conference, Washington, D.C.: The National Reading Conference, 1981, 113-136.

Durr, W. K., Windley, V., & Yates, M. C. Images. Boston: Houghton Mifflin, 1971, 1974.

Ginn & Company. 360 Reading Series. Boston: 1972.

Grimes, J. E. The thread of discourse. The Hague, Holland: Mouton, 1975.

Guszak, F. J. Teacher questioning and reading. The Reading Teacher, 1967, 21, 227-234.

Guthrie, J. T. Research views: Story comprehension. The Reading Teacher, 1977, 30, 574-75, 577.

Hansen, J. The effects of inference training and practice on young children's reading comprehension. Reading Research Quarterly, 1981, 16, 391-417.

Hansen, J., & Pearson, P. D. An instructional study: Improving the inferential comprehension of fourth grade good and poor readers. Journal of Educational Psychology, in press.

Johnston, P., & Pearson, P. D. Prior knowledge, connectivity, and the assessment of reading comprehension (Tech. Rep. No. 245). Urbana: University of Illinois, Center for the Study of Reading, June 1982.

Kintsch, W. Memory for prose. In C. N. Cofer (Ed.), The structure of human memory. W. H. Freeman & Co., 1975.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 1977, 9, 111-151.

Meyer, B. J. F. The organization of prose and its effects on memory. Amsterdam: North-Holland Publishing, 1975.

National Assessment of Educational Progress. Three National Assessments of Reading: Changes in Performance, 1979-80 (Report NO. 11-R-01). Denver: Education Commission of the States, 1981.

Nicholas, D. W., & Trabasso, T. Toward a taxonomy of inferences. In F. Wilkening & J. Becker (Eds.), Information integration by children. Hillsdale, N.J.: Erlbaum, in press.

Omanson, R. C., Warren, W. H., & Trabasso, T. Goals, themes, inferences and memory: A developmental study. Discourse Processing, 1978, 1, 337-354.

Paris, S. G. Integration and inference in children's comprehension and memory. In F. Restle, R. Shiffrin, J. Castellan, H. Lindman, & D. Pisoni (Eds.), Cognitive theory (Vol. 1). Hillsdale, N.J.: Erlbaum, 1975.

Paris, S. G., & Lindauer, B. K. The role of inference in children's comprehension and memory. Cognitive Psychology, 1976, 8, 217-227.

Paris, S. G., & Upton, L. R. Children's memory for inferential relationships in prose. Child Development, 1976, 47, 660-668.

Pearson, P. D., & Farstrup, A. Reading Performance in Minnesota. St. Paul, Minn.: State Department of Education: Minnesota Statewide Assessment, 1975.

Pearson, P. D., Hansen, J., & Gordon, C. The effect of background knowledge on young children's comprehension of explicit and implicit information. Journal of Reading Behavior, 1979, 11(3), 201-209.

Pearson, P. D., & Johnson, D. D. Teaching Reading Comprehension. New York: Holt, Rinehart & Winston, 1978.

Pichert, J., & Anderson, R. C. Taking different perspectives on a story. Journal of Educational Psychology, 1977, 69, 309-315.

Posner, M. I., & Snyder, C. R. R. Attention and cognition control. In R. Solso (Ed.), Information processing and cognition: The Loyola Symposium. Hillsdale, N.J.: Erlbaum, 1975.

Raphael, T. E., Winograd, P., & Pearson, P. D. Strategies children use when answering questions. In M. L. Kamil & A. J. Moe (Eds.), Perspectives on reading research and instruction. Washington, D.C.: National Reading Conference, Inc., 1980.

Rumelhart, D. E. Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance VI. London: Academic Press, 1977.

Rumelhart, D. E. Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Spiro, R. J. Remembering information from text: The 'state of schema' approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erblaum, 1977.

Steffensen, M. S., Joag-dev, C., & Anderson, R. C. A cross-cultural perspective on reading comprehension. Reading Research Quarterly, 1979, 15, 10-29.

Stein, N., & Glenn, C. G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), New directions in discourse processing. Norwood, N.J.: Ablex, 1979.

Stein, N. L., & Nezworski, T. The effects of organization and instructional set on story memory (Tech. Rep. No. 69). Urbana: University of Illinois, Center for the Study of Reading, March 1978.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.

Warren, W. H., Nicholas, D. W., & Trabasso, R. Event chains and inferences in understanding narratives. In R. O. Freedle (Ed.), Discourse processing: Multi-disciplinary approaches. Hillsdale, N.J.: Erlbaum, 1978.

Table 1

Instructional Activities for Development of Background Knowledge in Content and Structure Group

Story 1	Analogies
Story 2	Word Association
Story 3	Semantic Mapping
Story 4	Definition of Abstract Words Through Exemplification
Story 5	Word Association
Story 6	Semantic Mapping
Story 7	Analogies
Story 8	Semantic Mapping

Table 2

Developing Inferential Comprehension

Modeling and Feedback Technique

Objectives:

1. To make pupils aware of what the skill is by name and what the purpose of the skill is in any reading-thinking situation.

2. To make pupils aware of how to apply this skill in any material where it is needed.

In order to accomplish these two objectives the experimenter must teach what the skill is and its application.

Steps for teaching:

1. Define the skill:

   a) Name the skill (Making inferences)

   b) Give an example of the skill. (From Thorndyke, 1976) (i.e., When one reads "The hamburger chain owner was afraid his love for french fries would ruin his marriage," one might infer that the owner is fat.)

   c) Indicate how it is used in the reading-thinking situation.

   If the above text segment is followed by "He decided to join weight-watchers," the inference made is confirmed. However, if no explanatory statement follows, the student must access his/her stored knowledge and call up different scripts to make an appropriate inference such as 1) His wife didn't like the smell of french fries, or 2) His wife felt he spent too much time at his job.

2. Using suitable material from each basal reader selection, proceed with the following progression of steps:

   a) The experimenter provides an example: (Week 1) - Story 1

      1) The experimenter reads the material aloud while the pupils follow.

      2) The experimenter states the inference.

      3) The experimenter asks questions to elicit the clues from the story that support the inference.

Table 2 (continued)

and/or

The experimenter asks questions that access prior knowledge for omitted details or an explanation. These are clues to making the inference.

   b) The experimenter provides an example but asks for help: (Weeks 2 and 3)--Stories 2 and 3

      1) The experimenter reads the material.

      2) The experimenter states the inference.

      3) The experimenter asks the class for clues from the text that support the inference.

      and/or

      The experimenter asks the students to call up different scripts to make the inference.

   c) The experimenter provides help and asks for the inference: (Weeks 4 and 5)--Stories 4 and 5

      1) The experimenter reads the material.

      2) The experimenter states selected clues from the text that support the inference.

      and/or

      The experimenter calls up appropriate scripts to make the inference.

      3) The experimenter asks the students to state an inference that the textual clues or the scripts support.

   d) When the pupils demonstrate proficiency in steps a, b, c, above, proceed as follows: (Weeks 6 and 7)--Stories 6 and 7

      1) The experimenter reads the material.

      2) The experimenter asks the pupils to state an inference.

      3) The experimenter asks the pupils to state the clues from the text that support the inference.

Table 2 (continued)

and/or

to relate information from prior knowledge that has been used in the reasoning process.

e) Last week of research project: (Week 8)--Story 8

1) The experimenter reads the material.

2) The experimenter asks a question requiring the pupils to make an inference (these questions are actually worded as scriptally implicit and textually implicit questions during the final week).

3) The experimenter asks for the answer to the question

and

for information from the story (text) or from prior knowledge that supports the inference.

Table 3

Regression Effects: Comprehension as Measured by Instructional Worksheets

Step	Source of Variation	df	R^2 Change	F	p
1	MAT (M)	(1,40)	.32	19.09	<.001
2	Pretest (P)	(1,39)	.06	3.25	<.08
3	M x P	(1,38)	.03	2.35	<.13
4	C_1 $(E_1 + E_2 - 2C)$	(1,37)	.05	3.36	<.07
5	C_2 $(E_1 - E_2)$	(1,36)	.03	2.28	<.14
6	$C_1 \times P$	(1,34)	.01	.99	>.20
	$C_1 \times M$	(1,34)	.00	.16	>.20
	$C_2 \times P$	(1,32)	.05	3.97	<.07
	$C_2 \times M$	(1,32)	.00	.23	>.20
TOTAL EXPLAINED VARIANCE			.56		

Table 4

Means and Adjusted Means[a] for Questions

on the Instructional Worksheets

Question Type	Text Explicit		Text Implicit		Script Implicit		Total	
Treatment	*M*	*Ma*[a]	*M*	*Ma*[a]	*M*	*Ma*[a]	*M*	*Ma*[ab]
Content & Structure	5.93	7.45	4.15	3.92	2.86	1.39	4.31	4.25
Inference Awareness	5.93	7.35	4.79	5.07	4.21	3.96	4.97	5.47
Control	5.36	6.24	4.65	4.84	2.07	-.26	4.02	3.61
	4.74	7.01	4.53	4.61	3.07	1.69	4.23	4.44

Adjusted for pretest score on the Metropolitan Achievement Test.

These adjusted means correspond roughly to what the data for the groups looked like when the two contrasts were tested in the regression analysis.

Table 5

Regression Effects: Total Comprehension Scores

Summed Across the Four Post-Test Stories

Step	Source of Variation	df	R^2 Change	F	p
1	MAT (M)	(1,40)	.635	69.73	<.001
2	Pretest (P)	(1,39)	.056	7.02	<.02
3	M x P	(1,38)	.014	1.85	>.20
4	C_1 $(E_1 + E_2 - 2C)$	(1,37)	.045	6.60	<.02
5	C_2 $(E_1 - E_2)$	(1,36)	.019	2.95	<.09
6	$C_1 \times P$	(1,35)	.023	3.88	<.06
	$C_1 \times M$	(1,34)	.019	3.44	<.08
7	$C_2 \times P$	(1,33)	.012	2.23	<.15
	$C_2 \times M$				
TOTAL EXPLAINED VARIANCE			.823		

Separate Equations for Each Treatment: Pretest on Dependent Measure

Treatment	Intercept	B	SE_B	R^2
E_1 (Inference Awareness)	(y = 8.009 +	1.89 X)	.268	.81
E_2 (Content & Structure)	(y = .049 +	2.24 X)	.307	.81
C (Control)	(y =11.327 +	1.06 X)	.473	.29
$(E_1 + E_2)$	(y = 3.91 +	2.07 X)	.21	.81

Note. Simple R^2 for pretest on dependent measure for each treatment separately.

Table 6

Means and Adjusted Means for Story Comprehension

as a Function of Treatment and Question Type:

Sum for all 4 Stories

	Question Type							
	Text Explicit		Text Implicit		Script Implicit		Total	
Treatment	M	Ma[a]	M	Ma	M	Ma	M	Ma
Inference Awareness	12.85	20.83	12.42	23.35	6.78	1.22	10.69	15.14
Content & Structure	10.57	10.66	10.21	13.91	6.78	3.60	9.19	9.38
Control	11.00	7.35	9.35	9.15	5.00	-5.13	8.45	3.78
Total	11.47	12.95	10.68	15.46	6.19	- .11	9.44	9.44

[a] Adjusted for MAT and experimenter designed pretest.

Table 7

Summary of Separate Regressions for each Transfer Story:

($R^2 = R^2$ Change)

			Data Set							
			Immediate				Delayed			
			Text Access		No Access		Text Access		No Access	
Step	Variable	df	R^2	F	R^2	F	R^2	F	R^2	F
1	MAT	1,40	.5129**	42.11	.4826**	37.31	.4315**	30.37	.5633**	52.23
2	Pretest	1,39	.0153	1.27	.1294**	13.00	.0420	3.11	.0798**	8.80
3	Pretest x MAT	1,38	.0024	.20	.0110	1.11	.0459	3.63	.0041	.44
4	Prior Knowledge	1,37	.0006	.05	.0452*	5.03	.0484*	4.14	.0190	2.12
5	Contrast 1 (C1)	1,35	.0815**	7.40	.0083	.99	.0400+	3.85	.0121	1.43
	Contrast 2 (C2)	1,35	.0018	.16	.0307+	3.86	.0247	2.35	.0216	2.63
6	MAT X C1	1,33	.0016	.15	.0000	.00	.0101	.96	.0004	.04
	MAT x C2	1,33	.0254	2.30	.0157	1.86	.0120	1.52	.0024	.26
7	PT x C1	1,31	.0035	.30	.0073	.41	.0375+	3.82	.0016	.13
	PT x C2	1,31	.0002	.03	.0077	.91	.0037	.71	.0002	.02
8	PKT x C1	1,29	.0032	.42	.0058	2.30	.0000	.00	.0047	.48
	PKT x C2	1,29	.0127	1.09	.0411	5.54	.0004	.04	.0001	.00

** = $p < .01$

* = $p < .05$

+ = $.05 < p < .07$

Table 8

Regression Effects: Total Recall on Last Instructional Story

Step d	Source of Variation	df	R^2 Change	F	p
1	MATT (M)	(1,39)	.231	10.14	<.003
	Pretest Recall (P)	(1,39)	.129	7.85	<.008
2	C1 (E_2 - (C + E_1))	(1,37)	.184	15.13	<.001
	C2 (C - E_1)	(1,37)	.004	.00	>.20
3	M × C_1	(1,33)	.009	.81	>.20
	P × C_1	(1,33)	.047	4.25	<.05
	M × C_2	(1,33)	.012	1.30	>.20
	P × C_2	(1,33)	.000	.00	>.20
EXPLAINED VARIANCE			.618		

Separate Regressions

Source	Constant	B	SE_B	R^2	M	Ma
Content & Structure	(y = 3.68	+ 1.02X)	.29	.509	20.78	46.1
Inference Awareness	(y = 8.89	+ .292X)	.20	.154	12.93	-3.16
Control	(y =12.69	+ .121X)	.25	.019	14.50	5.27
(1A + C)	(y =10.61	+ .215X)	.15	.069	13.71	1.05

Note. Ma = Mean adjusted for MAT and first story recall

Table 9

Summary of Significant Regression Effects: Recall of Specific Categories of Information Using MAT score, Recall of the Same Category on the Initial Instructional Story, Treatment, and Their Interactions as Predictor Variables

Dependent Measure	Significant Effects	R^2 Change	F	Total
Major setting	None			.118
Minor setting	Pretest	.242**	8.87	.599**
	C_1[a]	.225**	16.05	
	MAT × C_1	.052*	3.92	
Initiating event	C_1	.116*	5.68	.304*
Internal Response	None			.068
Attempts	MAT	.255**	13.68	.474**
	Pretest	.083*	4.91	
	C_1	.078*	5.08	
Consequence	MAT	.160**	7.59	.363**
	Pretest	.096*	5.02	
	C_1	.079*	4.54	
Reaction	C_1	.095*	4.63	.304*
Resolution	C_1	.213**	10.36	.306*

* $p < .05$

** $p < .01$

[a] C_1 = Contrast 1 = (Content & Structure) versus (Inference Awareness and Control); C_2, which contrasted Inference Awareness with Control, never proved significant.

Table 10

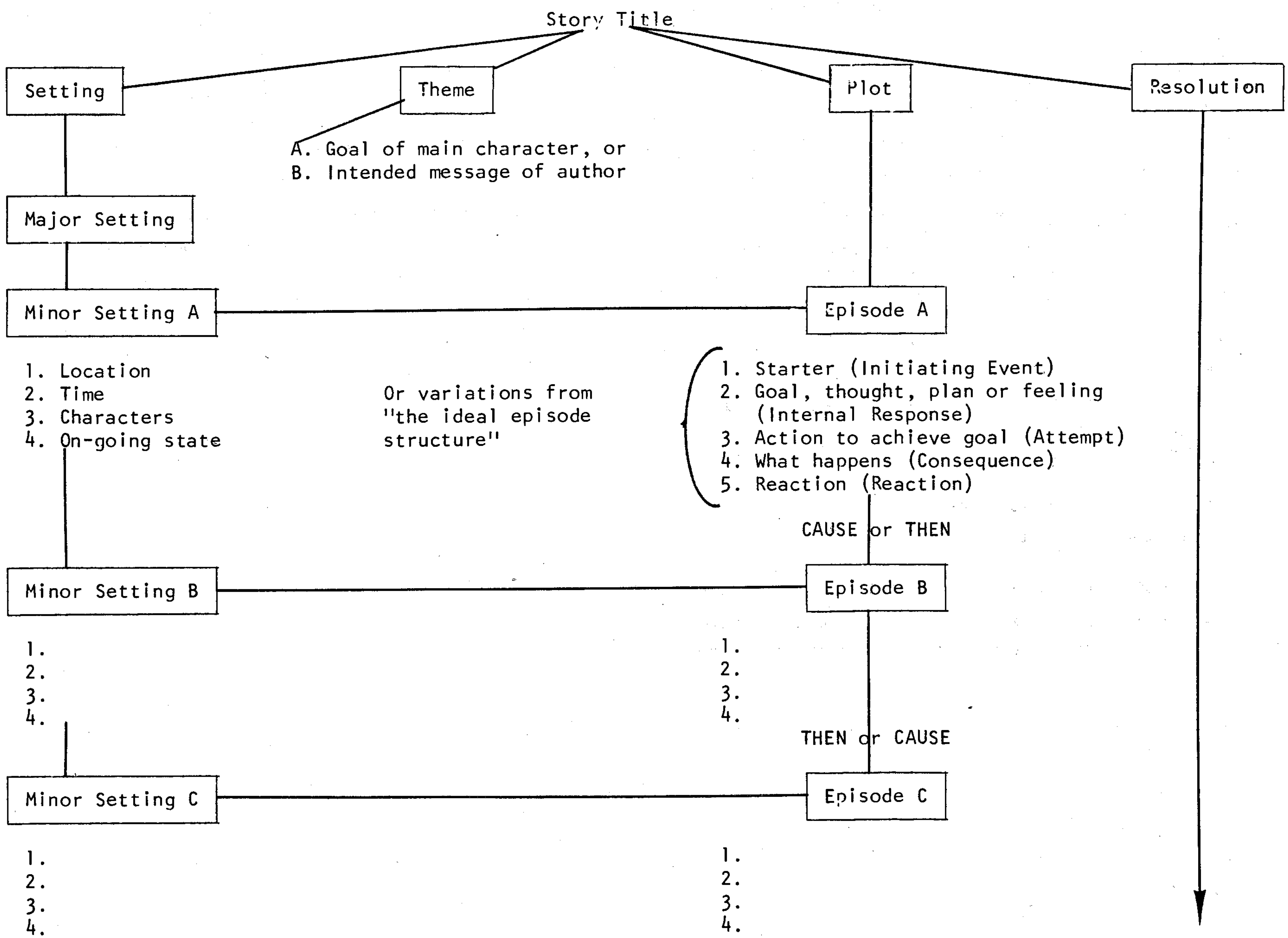
Regression Effects: Sum of Metropolitan Achievement Test Scores (vocabulary and comprehension) administered as a post-test using pretest MAT scores and treatment as predictors

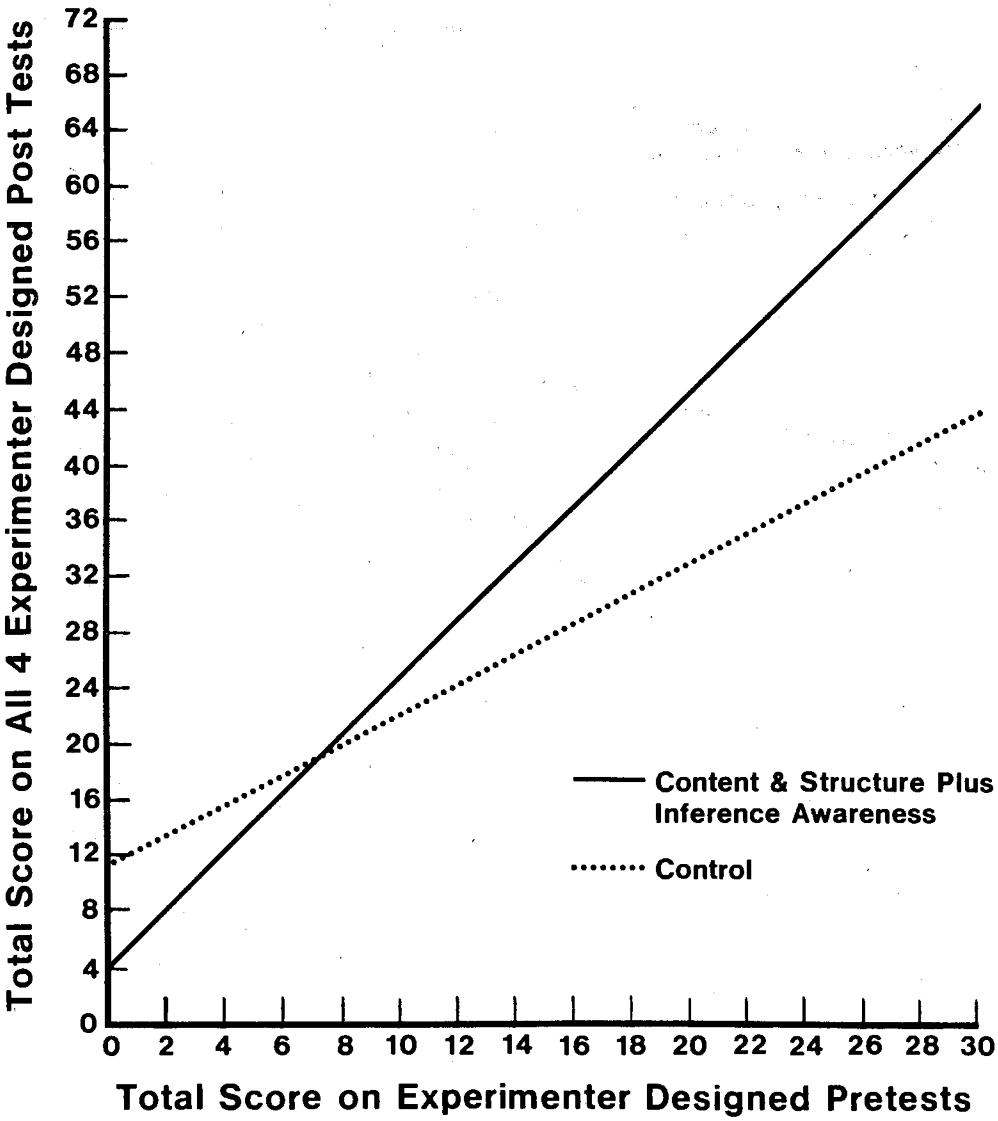
Step	Source	R^2_a	F	df	p
1	MAT 1T	.73	108.99	(1,40)	<.001
2	C_1	.001	.18	(1,38)	>.05
	C_2	<.001	.07	(1,38)	>.05
3	C_1 MAT 1 T	.03	4.15	(1,36)	<.05
	C_2 MAT 1 T	<.001	.06	(1,36)	>.25
	TOTAL EXPLAINED VARIANCE	.76			

Separate Regressions

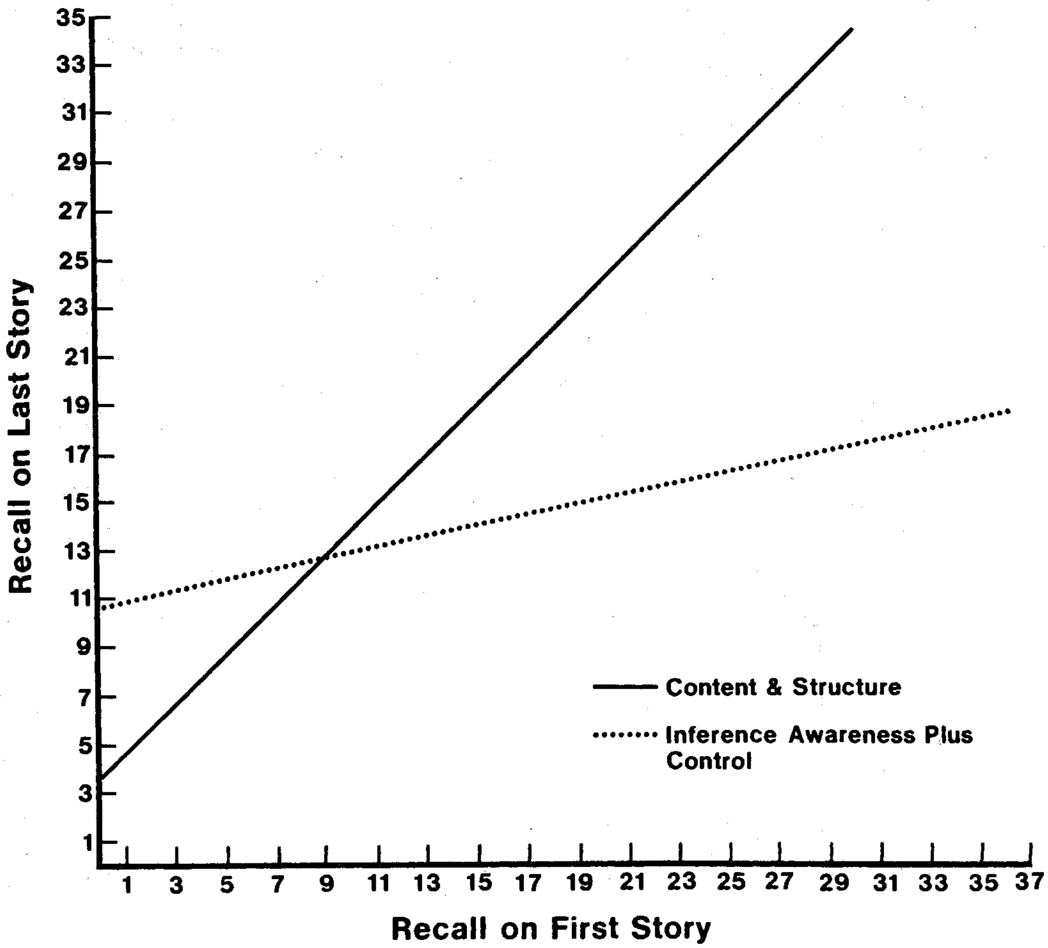
Source		Constant		B	SE_B	R^2	M	sd
1A + CS	Y =	(12.95	+	.85X)	.07	.90	63.42	13.62
Control	Y =	(32.04	+	.51X)	.18	.40	62.71	10.28

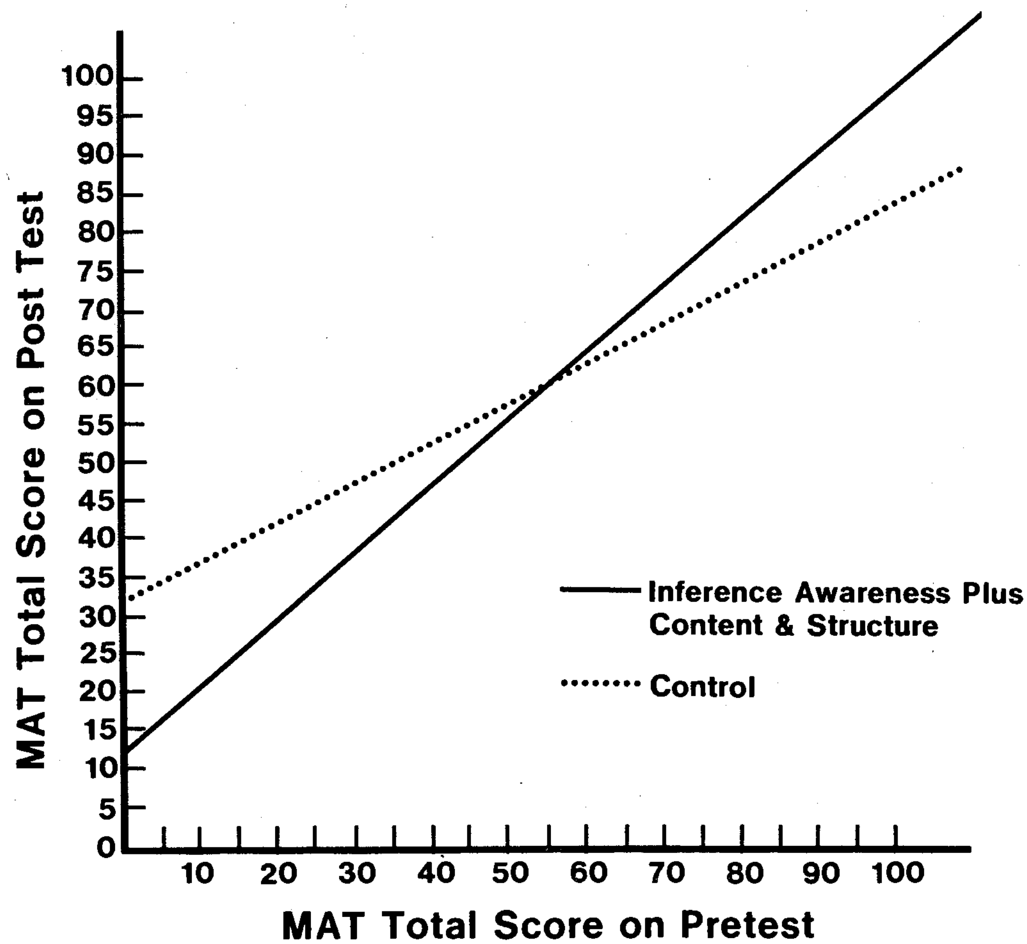
Figure Captions


Story Title
Setting
Theme
Plot
Resolution
A. Goal of main character, or
B. Intended message of author
Major Setting
Minor Setting A
Episode A
1. Location
2. Time
3. Characters
4. On-going state
Or variations from "the ideal episode structure"
1. Starter (Initiating Event)
2. Goal, thought, plan or feeling (Internal Response)
3. Action to achieve goal (Attempt)
4. What happens (Consequence)
5. Reaction (Reaction)
CAUSE or THEN
Minor Setting B
Episode B
1.
2.
3.
4.
1.
2.
3.
4.
THEN or CAUSE
Minor Setting C
Episode C
1.
2.
3.
4.
1.
2.
3.
4.



Total Score on All 4 Experimenter Designed Post Tests
72
68
64
60
56
52
48
44
40
36
32
28
24
20
16
12
8
4
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Total Score on Experimenter Designed Pretests
Content & Structure Plus Inference Awareness
Control


35
33
31
29
27
25
23
21
19
17
15
13
11
9
7
5
3
1
Recall on Last Story
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37
Recall on First Story
Content & Structure
Inference Awareness Plus Control


MAT Total Score on Post Test
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
10
20
30
40
50
60
70
80
90
100
MAT Total Score on Pretest
Inference Awareness Plus Content & Structure
Control

This page is intentionally blank.