American Educational Research Journal
Summer 1985, Vol. 22, No. 2, Pp. 217–235

Increasing Students' Awareness of Sources of Information for Answering Questions

TAFFY E. RAPHAEL
Michigan State University

and

P. DAVID PEARSON
University of Illinois

The ubiquity of oral and written questions in classrooms and reading tasks has been well established. It is obvious that students' background knowledge is a primary determinant of how well they can answer comprehension-promoting or assessment-oriented questions. Less obvious is the need to access appropriate sources of information used in answering questions. A training study was conducted with 59 sixth grade students of high, average, and low reading ability levels, instructing half of students in each ability level in the relationship of a question, the text to which the question refers, and the reader's knowledge base. Repeated measures multivariate analyses of variance on two passages indicate that the training enhanced students' understanding of task demands of questions and their answer quality as measured by (a) their ability to correctly identify the question-answer relationship represented, (b) their consistency in providing an answer from the source of information they had indicated they were using, and (c) the quality of their answers. The training appeared to be most facilitative for those of average and lower ability levels, consistent with similar findings in other studies that trained cognitive skills. Knowledge of sources of information may be fundamental to students' ability to access appropriate information for answering comprehensive questions.

Questions are a well-established way of life in classroom reading instruction (Bartolome, 1969; Durkin, 1978–79; Guszak, 1967; Hare, 1982). They are used to promote comprehension evidenced by, for example, the long line of research on the use of questions as adjunct aids (e.g., Anderson & Biddle, 1975; Rickards & Hatcher, 1978; Rothkopf, 1965, 1966). Further, it is often on the basis of students' ability to answer questions on either formal (i.e., standardized tests) or informal (i.e., informal reading inventories or daily assignments) tasks that we label students as either good or poor readers.

However, what much of the research on questions has failed to consider is the active role of the learners (Baker & Brown, 1984; Whimbey, 1975) in becoming more facile and flexible in the way they approach question-answering tasks. In particular, instructional research on questioning behavior has not considered the fact that different questioning tasks impose not only different cognitive reasoning demands, but also differential attention to the potential sources of information students can, could, or should use in answering questions. Indeed, most taxonomies of questions (e.g., Barrett, 1976; Bloom, 1956; Sanders, 1966) deal only with the presumed cognitive demand of the question, and fail to acknowledge the fact that the availability (or lack of availability) of information can render what appears to be a cognitively low level task difficult, or reduce a presumably high level task to little more than a question/sentence matching task (see Pearson & Johnson, 1978).

Recent reconceptualizations of the cognitive demands of questions (Pearson, 1983; Pearson & Johnson, 1978) as they relate to the information explicitly available during reading (i.e., in the text) as well as implicitly available (i.e., in the reader's store of prior knowledge) have provided a framework for studying the development of question-answering behavior during reading tasks. The Pearson and Johnson (1978) taxonomy is unique in that it does not classify questions in isolation; instead it emphasizes the three-way relationship among question, text, and learner.

The Pearson and Johnson (1978) taxonomy of questions has been used to examine students' knowledge and use of question-answering strategies in two descriptive studies. Raphael, Winograd, and Pearson (1980) found that the ability to recognize appropriate information sources appeared to be related to the quality of answers students provided to questions. That is, the more successful students (i.e., those with answers that were both correct and complete) tended to be flexible in their use of information sources for providing these answers. They tended to use explicitly stated information when a low level literal question was asked, to integrate text information when the question required such integration, and to use information from prior knowledge when the text did not provide an explicit answer to an inference question. In contrast, poor readers were apparently

unaware of, or at least often ignored, such variations in task demands of questions; that is, they tended to use a strategy quite independent of what was appropriate given the requirements of the question and the availability of information in the text.

To understand further the relationship between knowledge of information sources and ability to answer comprehension questions, Wonnacott and Raphael (1982) examined the relationships between third and sixth grade students' knowledge of the question-answering process and their performance on the three types of comprehension question. They used a combination of measures, including data from interviews, observations, and performance on question-answering tasks. Results indicated that metacognitive insight, as measured by students' ability to verbalize their understanding of the process of question answering, was a strong positive correlate of performance on comprehension questions.

Although the studies by Raphael et al. (1980) and by Wonnacott and Raphael (1982) suggested that knowledge about the question-answering process and sources of information for answering comprehension questions is important, both studies were essentially descriptive. Thus, they cannot provide causal explanations of the relationship between students' strategic (i.e., metacognitive) knowledge and actual performance. Belmont and Butterfield (1977) suggested that training studies can provide such information about cognitive processes. They proposed that successful intervention implies a causal relationship between the means trained and the goal to be reached; that is, one can learn if a component of a process is related to a goal, or cognitive outcome, by manipulating the process. Similar suggestions were proposed by Brown, Campione, and Day (1981) in their discussion of "informed" training studies where students are taught about a strategy and induced to use it and are given some indication of the significance of the strategy. Finally, Sternberg (1981) provided an extensive discussion of prerequisites for general programs that attempt to train cognitive skills, including suggestions such as the need to link such training to "real-world behavior" as well as to theoretical issues.

Accordingly, the specific purpose of the present training study was to examine the role of knowledge of information sources in children's question-answering abilities through the examination of an instructional program designed to heighten their awareness of information sources. It was predicted that as a result of training, (a) students' awareness of appropriate sources of information for answering comprehension questions would be heightened, (b) students' strategies for providing answer information would be consistent with their identification of question-answer relationships, and (c) the quality of their answers would improve. Finally, it was predicted that these outcomes would vary with the students' reading levels, given the differential performance of students of varying levels in both the Raphael et al. (1980) and the Wonnacott and Raphael (1982) studies.

METHOD

Subjects and Design

Fifty-nine sixth grade students, from three classrooms in a suburban school district of a large midwestern city, were blocked on ability using a combination of teacher judgment, reading group membership, and standardizing reading comprehension test scores from the Stanford Achievement Test. Students who were labeled reading disabled or students who had decoding problems such that they would not be able to read the materials had been previously eliminated from the larger subject pool. The remaining 59 students were then randomly assigned to treatment and control groups. There were 10 students in all cells except the low ability control group, which contained 9.

A $3 \times 2 \times 3$ randomized block design was used with between-subjects factors of ability (high, average, low) and treatment (trained, control), and the within-subjects factor of question-answer relationship (text explicit, text implicit, script implicit).

Materials

All questions for materials and the training procedures themselves were adapted from the Pearson and Johnson (1978) taxonomy of question types. This categorization scheme was selected because it is unique in the method by which a question is classified. While the majority of taxonomies assume that questions can be classified in isolation, this three-category taxonomy underscores the necessity of identifying questions according to their relationship to two primary sources of information: the text to which the question refers and the knowledge base of the reader. Thus rather than speak of question types, with this taxonomy it is more appropriate to refer to question-answer relationships (QARs). A text explicit (TE) question-answer relationship is a question with words comprising both the question and answer information stated explicitly in a single sentence of the text. A text implicit (TI) question-answer relationship is a question with answer information available in the text, but requiring the reader to integrate information across sentences or paragraphs in the text. A script implicit (SI) question-answer relationship is a question for which the information appropriate as an answer is not available in the text, requiring the readers to fill in the gaps from their own knowledge bases.

The three QARs—text explicit, text implicit, and script implicit—were explained to students using the following mnemonics, respectively: (1) Right There meant that words used to create the question and words used for the answer are "right there" in the same sentence. (2) Think and Search meant that the answer is in the text, but words used to create the question and those used for an appropriate answer would be found in two or more sentences; you would have to "think and search" for an answer across

sentences and paragraphs. (3) On My Own meant that the answer is not found in the text; rather, you would think to yourself that "I have to find this answer 'on my own.'"

Instructional materials. Three booklets were developed for use in the training aspect of the study. The first booklet was designed to introduce the concept of QARs to the students using both text and visual mnemonics (e.g., a drawing of an opened book and a cartoon of a child's head for Think and Search QARs). In this booklet, a series of brief passages (two to five sentences) were used, each followed by one question from each of the three QAR categories. The following story and question set is an example of the brief passages in the students' first booklet:

> Dennis sat in an old wood rocking chair. He rocked harder and harder. Suddenly he found himself sitting on the floor!
>
> *Right There:* What kind of chair did Dennis sit in? (old wood rocking chair)
> *Think and Search:* What did Dennis do while sitting in the chair? (rocked harder and harder)
> *On My Own:* Why did Dennis find himself sitting on the floor? (rocked so hard the chair tipped over)

Of course, such brief paragraphs limit the amount of integration required for think and search questions. Truer integration of information was possible only with the longer passages; these brief passages indicated that more than one sentence of a passage must be used. In the example, the question was derived from the one sentence, while the answer was found in another. In this case, the integration was simple: "he" refers to "Dennis." Based on the principles of shaping and fading, or moving from providing a great deal to very little support, the tasks began with the researcher and the booklet providing text, question, answer, QAR label, and a discussion of the reasons why a given label was appropriate. Next, text, question, answer, and QAR label were given, with the students providing the explanation and receiving immediate feedback in their groups from the instructor on the accuracy and completeness of their explanations. Third, text, question, and answer were provided, with students supplying the labeling and justification. Finally, students were provided with text and questions and were asked both to answer the questions and to indicate the QAR represented, using the following format, referred to as the QAR Task:

1. Why was brush popping a dangerous activity for the cowboy?
   Right There _____
   Think and Search _____
   On My Own _____

Students answered the question on the line next to the QAR they felt the question and answer best represented. The materials in the first booklet

reflect this progression. In the second and third booklets, students practiced applying their QAR knowledge on progressively longer expository texts, gradually building from 75-word to 400-word passages selected from naturally occurring classroom materials; each passage was followed by questions with the QAR task. An example of one of the texts from the second booklet follows:

> Some musical instruments can make high tones, while others make low tones. These sounds are made by vibrations in the air. With a piano or violin, it is easy to see that the vibrating of the strings makes the sounds that come from these instruments. It is not so easy to see what vibrates in a horn to produce the sounds.
>
> Some horns have a reed that vibrates when a person blows into the instrument. The vibrating reed causes the air inside the instrument to vibrate as well, and that is what creates the sound. By pressing down some keys on the horn to make the space taken up by the vibrating air shorter or longer, a player can make low or high tones.
>
> Some horns do not have reeds in them. In playing these horns, it is the lips of the player which vibrate instead of the reed.
>
> There are many different instruments in a band. They all play together most of the time. If the sound waves do not blend together, an unpleasant noise results. It takes a great deal of practice to learn to produce sound waves that blend together well.
>
> *Right There:* What happens if sound waves do not blend together well? (an unpleasant noise results)
> *Think and Search:* What are three different ways that instruments can produce sound waves? (by vibrating strings, reeds, or players' lips to make the air vibrate)
> *On My Own:* What causes sound waves to be made when a drum is played? (the drummer's sticks make the cover of the drum vibrate, and that sends out sound waves)

The total number of questions per QAR category was balanced across stories; students were exposed to each QAR an equal number of times in each booklet.

Test materials. Four passages on familiar topics were developed. Each student read two passages. One passage, "Cowboys of the Old West," of approximately low sixth grade readability (about 550 words in length) was read by all students participating in the study. A second set of three passages formed the level-appropriate passage set. Students of low reading ability read a passage, "About Dogs," at approximately the fourth grade level (about 400 words), students of high ability read "The Zoo Story," at the eighth grade level (approximately 700 words), and the average students read a second "filler" (4th grade level, approximately 450 words) included only to insure an even amount of time spent participating in the study. This passage was not included in the analyses, since for average students the same passage (6th grade level Cowboy passage) was used in both

analyses. Readability was determined using the Fry (1968) readability formula. For each passage, there were 18 corresponding questions, 6 from each QAR category. All passages and questions had been piloted; these texts were determined to be of essentially equivalent topic familiarity and interest, and questions were judged by skilled adult readers to belong to the category for which they were intended.

Procedure

Instructional. Students in the instructional group received 4 days of instruction, each session approximately 40 minutes long. On the first day, students were introduced to QARs in groups, with the researcher providing definitions and visual representations to make clear the three QAR concepts. Students were then led through the phases of the first booklet as described earlier. Feedback was provided on a group or individual basis, depending on the activity, focusing on accuracy of students' explanations, their ability to recognize the correct QAR when provided with a question and answer, the completeness and accuracy of their answers when provided with only the question, and the consistency between their QAR selection and the source of information for their answer (e.g., if an On My Own QAR was selected, their answer was from their knowledge base and appropriately answered the question). On the second day, students worked through the second booklet, which consisted of longer passages and more questions per passage. The instructor read aloud the first passage and answered questions in a group format. Students then worked through the remainder of the booklet passages and questions, while feedback on QAR selection and answer accuracy was provided individually. On the third day, students practiced with a full-length passage divided into four segments, each followed by six questions (two from each QAR category), equivalent in length to a typical basal reader story or section of a content area chapter. For the first segment, students read the segment and corresponding questions, identified the questions by QAR category, and answered them working individually. Then they corrected the QARs together in the larger group. The last three segments were completed individually and served as a criterion test. The fourth day had been planned for individually tutoring those who had not reached a 75% accuracy level in QAR identifications. This was not necessary, however, since all students were successful. The fifth day involved the experimental test, which lasted approximately 1 hour.

The control group students did not participate until the time of testing. Although it may be argued that one type of control group would involve practice answering questions similar to those taught in the instructional group, we elected not to do this for two reasons. First, studies (e.g., Tierney, 1976) have shown that students spend a large proportion of their school

day answering questions, thus it is a very familiar and frequent task. Second, other studies (e.g., Raphael & Wonnacott, in press) included a control group that practiced using all passages in the training booklet (but without reference to question-answer relationships) and a no-treatment control group. There were no significant differences between these groups. Thus, we elected to use the simpler control in this study.

Testing. Students in both instructional and control conditions were tested on the same day, each responding to the common passage with its related 18 questions and to one passage and question set from the level-appropriate passage set. Stories were counterbalanced across all subjects. For trained students, the only directions were to read the passage and answer the questions, writing answers on the blank next to the QAR represented. Control group students received an explanation of definitions of the three QARs (similar to the first day's explanation for the instructional group), were reminded that some of the answers would not be found in their texts, and then practiced identifying QARs on two brief texts with related questions. Then they were asked to read the passages and answer the questions on the blank next to the appropriate QAR. This introduction, used in pilot studies, had been found to be sufficient for skilled adult readers to complete the task with 98% accuracy and was considered to be the minimum necessary to provide an adequate control group still capable of using the QAR task.

Scoring

Scores were created independently for both the QAR identifications and for actual answers to the questions. Data from pilot studies were used to establish the criteria for both QAR identifications and for question-answer quality. For the QAR identifications in the pilot studies, 95% of skilled readers responding to a given question had to agree on its QAR classification for it to be included in the materials for this study. To receive credit, students in the present study had to identify the QAR as belonging to the same category as had been previously established by the pilot data. The total number of correct QAR identifications formed the first dependent measure, *hits.* The second set of criteria for which students' data were scored concerned their actual answers to the comprehension questions, their *response quality.* Answers were scored according to two factors: (a) correct versus incorrect, and (b) knowledge based versus text based. Thus one could have correct or incorrect text-based answers and correct or incorrect knowledge-based answers. Percent agreement for rating answers to the comprehension questions between two judges was .97. This measure indicated the total number of correct answers. The third dependent measure, *matches*, was created from the QAR identifications and the answers to the questions. To create this measure a matrix was established whereby

students received a point for each case in which they had a correct or incorrect text-based answer for any question that they had labeled (correctly or not) as a Right There or as a Think and Search. Similarly, credit was given for each case in which an answer from their knowledge base (correct or not) was provided to any question they had labeled as an On My Own. The total number possible for each QAR, regardless of dependent measure, was six.

RESULTS

Results are reported and interpreted in terms of the three dependent measures. The first, hits, was a measure of the students' sensitivity to the task demands of a question. That is, it was concerned with the question, Were the students able to identify the question-answer relationship correctly? The second, response quality, was a measure of the degree to which teaching students about information sources for answering comprehension questions enhanced the quality of students' answers. That is, it concerned the question, How accurate and complete were the students' answers? The third, matches, was an indication of the degree to which students had internalized the knowledge about QARs, since it considered the question, Did the students do what they indicated they should or would do?

Two $3 \times 2 \times 3$ repeated measures MANOVAs are reported, one each for the common passage and the level-appropriate passage set using the three dependent measures. The between-subjects factors were ability and condition, and the within-subjects factor of QAR was divided into two a priori contrasts, using Helmert's contrasts in the multivariate procedure described in Bock (1975). The contrasts were determined based on previous research (e.g., Hansen & Pearson, 1983; Lipson, 1983; Pearson, Hansen, & Gordon, 1979) that suggested that questions with answers explicitly stated in the text are more easily answered by students than are questions with answers implied by texts. Further, students perform differentially on text-based and knowledge-based inference questions (Wixson, 1984). Thus, the first contrast was between performances on explicit and implicit QARs by comparing performance levels on text explicit with the average of text implicit and script implicit scores. This is identified as QAR(1). The second was a contrast between performances on text-implicit and script-implicit QARs, identified as QAR(2). The Wilk's lambda multivariate test of significance was used, followed by univariate F tests where appropriate. Since theoretical research on performance in question answering tasks, knowledge about question answering processes, and the causal relationship between the two is not yet well developed, a causal model has not been postulated. Therefore, simple univariate rather than step-down univariate tests were used. All means and univariate F tests for each of the three main effects referred to throughout the report of results are listed for both

passages within each dependent measure—hits, response quality, and matches—in Tables I, II, and III, respectively.

Common Passage

A three-way, multivariate, repeated measures analysis of variance was used to analyze the data for the three dependent measures of hits, response quality, and matches on the common passage. The results indicated significant multivariate effects for ability, $F(6,102) = 6.21$, $p < .01$; condition, $F(3,51) = 6.17$, $p < .01$; and QAR(2), $F(3,51) = 25.83$, $p < .01$, all of which were involved in a significant three-way interaction, $F(6,102) = 3.86$, $p < .01$. The multivariate difference for QAR(1) could be accounted for by chance alone, $F(3,51) = 2.20$, $p > .05$, as was the case for all two-way interactions. The follow-up examinations were performed, first on the significant interaction and then on the significant main effects.

The significant effect for the ability × condition × QAR(2) interaction could be attributed solely to differences on the measure of response quality, $F(2,52) = 8.77$, $p < .01$ (see Figure 1). Performance of students who had received training, while superior in general to that of the control group students, was most pronounced on text implicit questions among low ability students; training did not help the lower ability students in their performance on knowledge-based questions.

Follow-up univariate F tests for ability revealed significant differences

TABLE I

Table of Means and F Values for the Hits Dependent Measures for Common Passage (CP) and Level-Appropriate Passage Set (LAPS) Comparisons

	CP	LAPS
Ability	$F(2,53) = 10.36$**	$F(2,53) = 10.16$**
High	4.26	4.44
Average	3.60	3.60
Low	3.00	3.42
Condition	$F(1,53) = 12.56$**	$F(1,53) = 3.99$*
Trained	4.02	4.08
Control	3.00	3.60
QAR	a$F(1,53) = 1.15$	a$F(1,53) = 4.60$*
	b$F(1,53) = .10$	b$F(1,53) = 5.60$*
Text Explicit	3.45	4.14
Text Implicit	3.60	3.36
Script Implicit	3.53	3.96

Note. Contrasts: aQAR(1) = explicit versus implicit; bQAR(2) = text versus script.
* $p < .05$.
** $p < .01$.

TABLE II

Table of Means and F Values for the Response Quality Dependent Measure for Common Passage (CP) and Level Appropriate Passage Set (LAPS) Comparison

	CP	LAPS
Ability	$F(2,53) = 7.55^{**}$	$F(2,53) = 3.56^{*}$
High	5.22	5.34
Average	5.04	5.04
Low	4.26	4.68
Condition	$F(1,53) = 13.60^{*}$	$F(1,53) = 5.11^{*}$
Trained	5.28	5.16
Control	4.44	4.80
QAR	$aF(1,53) = 6.73^{**}$	$aF(1,53) = 10.54^{**}$
	$bF(1,53) = 28.46^{**}$	$bF(1,53) = 15.35^{**}$
Text Explicit	5.10	5.28
Text Implicit	5.22	5.16
Script Implicit	4.26	4.50

Note. Contrasts: aQAR(1) = explicit versus implicit; bQAR(2) = text versus script.
$^{*}p < .05$.
$^{**}p < .01$.

TABLE III

Table of Means and F Values for the Matches Dependent Measure for the Common Passage (CP) and Level-Appropriate Passage Set (LAPS) Comparisons

	CP	LAPS
Ability	$F(2,53) = 19.38^{**}$	$F(2,53) = 4.22^{*}$
High	5.46	5.22
Average	4.62	4.62
Low	3.60	4.74
Condition	$F(1,53) = 3.99^{*}$	$F(1,53) = 6.21^{*}$
Trained	4.80	5.10
Control	4.32	4.62
QAR	$aF(1,53) = .26$	$aF(1,53) = 11.24^{**}$
	$bF(1,53) = 75.79^{**}$	$bF(1,53) = 31.21^{**}$
Text Explicit	4.62	5.16
Text Implicit	5.10	5.16
Script Implicit	3.96	4.26

Note. Contrasts: aQAR(1) = explicit versus implicit; bQAR(2) = text versus script.
$^{*}p < .05$.
$^{**}p < .01$.

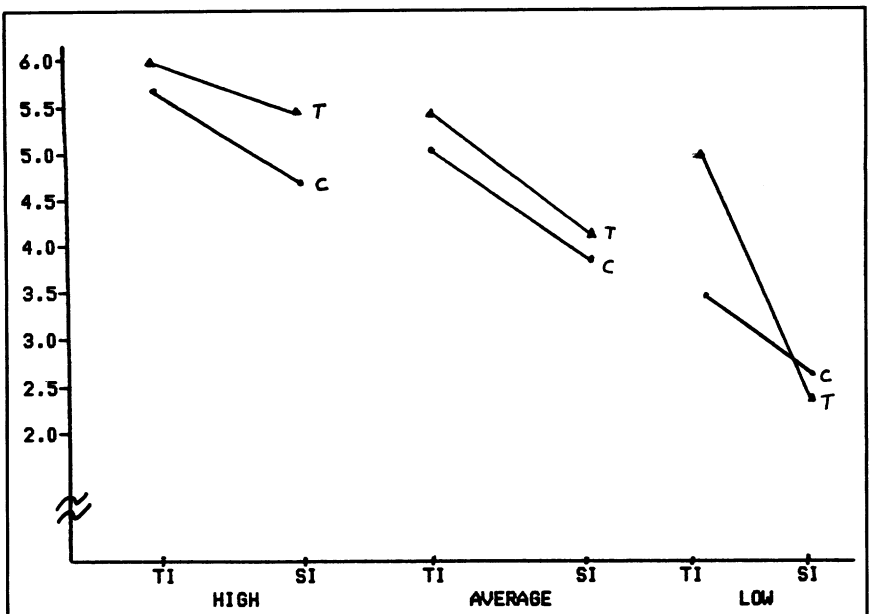

FIGURE 1. Ability × Condition × QAR(2) interaction on the Response Quality dependent measure.

on the three dependent measures. Since the measure of response quality was involved in the higher order interaction, it will not be discussed as a main effect. Differences for both the hits and the matches measures were in predicted directions. Students of high ability demonstrated more accuracy in QAR identification than those of average ability (see Table I). Average students performed with more accuracy than low ability students. A similar pattern can be seen in terms of consistency between QAR identifications and the location of answer information by examining the means for matches in Table III. Again, high ability students were more successful than average ability students, who in turn were more successful than low ability students.

The multivariate effect for condition was attributed to differences in performance on hits, response quality, and matches. These were in predicted directions, as seen in Tables I, II, and III, with students in the group receiving instruction performing at a higher level than those students in the control group on all measures.

The multivariate effect for QAR(2) was attributed to differences in both response quality and matches across text and script categories. The difference in response quality can be explained in terms of the higher order interaction with ability and condition discussed previously. The differences

in performance on the match dependent measure, as seen in Table III, indicate that students demonstrated more consistency between their QAR identifications and answer information source used for text than for script-based questions.

Level-Appropriate Passage Set

A three-way, multivariate, repeated measures analysis of variance was also used to examine data for hits, response quality, and matches for the level-appropriate passage set. The results indicated three significant multivariate effects. These were for ability, $F(6,104) = 5.25$, $p < .01$; for QAR(1), $F(3,51) = 5.09$, $p < .01$; and for QAR(2), $F(3,51) = 13.51$, $p < .01$. The multivariate effect for condition did not reach significance, $F(3,51) = 2.51$, $p = .07$. There were no significant interactions.

The multivariate effect for ability was attributable to significant differences on all three dependent measures. Differences (as seen in Tables I, II, and III) on these measures were in predicted directions, with high ability students performing at a level superior to average students, who were superior in performance to low ability students.

Although the multivariate effect for condition did not reach a traditional significance level ($p = .07$), it was examined further in an attempt to understand the unexpected lack of significant effect. Significant univariate effects for the three dependent measures were revealed. Tables I, II, and III indicate that the differences were in the predicted direction on the three measures, with trained students performing at a higher level than those in the control group. This pattern was consistent with findings on the common passage analyses, but can be considered only as trends in the data.

DISCUSSION

Three predictions made at the outset of the study concerned the efficacy of a program for heightening students' awareness of and sensitivity to sources of information for answering comprehension questions. These predictions are considered in terms of both the common passage and the level-appropriate passage set. The common passage comparison was intended to reflect the situation students are in through most of their academic and nonacademic experiences. In academic settings, students are asked to read materials (e.g., science or social studies texts) at their expected grade level, regardless of their specific reading ability levels. In nonacademic settings, students may read newspapers, magazines, and books that are not tailored to their specific reading level. The level-appropriate passage set was intended to reflect the more controlled situation in the developmental reading program in which students are assigned to a particular basal reader based on their reading ability level. Although both comparisons have ecological validity, the common passage is the more prevalent

situation. The results are discussed across both settings, with differences in patterns noted where appropriate.

First, it was predicted that instruction in question and answer relationships would enhance students' sensitivity to task demands of questions, as indicated by their ability to judge the source of information most likely to be appropriate for answering the question. This is precisely what the data suggest from follow-up tests to both MANOVAs. Instruction enhanced students' awareness of task demands as indicated by their recognition of the QAR category to which a question belonged. This was statistically confirmed in the situation using grade level materials, and the trend was apparent when students read passages at their specific reading levels. This sensitivity was also a function of ability in that accuracy in QAR identifications increased as students increased in reading ability. Finally, consistent with the research literature (Au & Ignacio, 1983; Wixson, 1984; Wonnacott, 1983), text implicit QARs were the most difficult to identify when reading materials of appropriate reading difficulty. These text implicit QARs require an awareness of the interplay between the readers' own knowledge base and how to use that knowledge to integrate text information appropriately, a more sophisticated skill than deciding whether information is present in text.

The second prediction was that instruction would improve the quality of the students' answers. This prediction was confirmed on the common passage comparison, and though not at a traditional significance level, the trained students also performed at a higher level of the level-appropriate passage set. Research (e.g., Day, 1980; Hansen & Pearson, 1983) has suggested that when explicitly teaching students to use effective reading-related strategies, training effects are most notable with those students of lower ability levels. This also is intuitively sensible in that students of higher ability levels presumably already apply their knowledge of strategies and tasks successfully, and often are labeled as such because of their performance on question answering tasks on standardized and teacher-constructed tests. Lower ability students were helped most on text-based, in contrast to the knowledge-based, questions. The QAR instruction they received may have provided a foundation for using strategies to access information from text. Thus, the trained students of average ability performed at the level of high ability control group students in terms of the quality of their answers; trained low ability students performed at the level of average control group students on text-based questions. Yet, with script implicit questions, merely telling the lower ability readers to go to their heads was not sufficient. Students of average and lower ability levels needed assistance on actually accessing appropriate information, or as one student stated in response to a script implicit QAR, "I went to my head, but there wasn't anything there!" Research in the application of background knowl-

edge by Hansen and Pearson (1983) and Gordon (1979) has demonstrated improvements in reading comprehension after assisting students to apply relevant background knowledge, complementing the findings in this study.

A third prediction was that as a result of the training, students would demonstrate greater consistency between their indication of QAR category and the invited source of information used in answering the specific question. Trained students demonstrated superior performance on both passages read; however, differences were significant only on the common passage. Consistency was also a function of ability level on both passages, with high ability students demonstrating more consistency than average students, who in turn were more consistent than low ability students.

One finding of interest concerned the use of both the common passage and the level-appropriate passage set. In designing the study, we debated two methodological arguments regarding materials to be used when drawing ability comparisons. Some (e.g., Taylor, 1980) have argued that content can be held constant, but that actual readability of the passage should be at the reading level of the student. Others (e.g., Tierney, Bridge, & Cera, 1978–79) have cautioned that it may be unsound to draw comparisons across different passages, that this confounds any effect with materials. Both arguments have merit, thus the methodological, as well as the ecological value of using the two different passages for each student. On the measure of hits—which required only that students recognize whether answer information was available in the text or not, or whether that information was in a single sentence (but did not require that they understand the information well)—there was little difference across passages in the amount of variance attributed to reading ability. Correlations between reading comprehension, as measured by standardized test scores, and ability to recognize task demands (hits) on the common passage (r = .49) and on the level-appropriate passage set (r = .48) did not differ. The measure of matches, which required that students provide answer information from the source they had indicated, but did not require the accuracy of that answer, correlated slightly more with ability on the common passage (r = .41) than on the level-appropriate passage set (r = .35). Finally, with the measure that required the most reliance on reading ability—the response quality—much more of the variance was accounted for by ability on the common (r = .68) than on the level-appropriate passage set (r = .32), which attempted to equalize for ability. It appears that there is no simple solution to the problem of materials to be used in cross-ability studies. Reading level passages confound content with ability level; with common passages students of differing ability will experience different levels of difficulty. Perhaps being conservative in interpreting results is the only solution until effects are replicated across a number of passages.

Research in the general area of metacognition suggests that students can

be made more aware of both declarative and procedural knowledge related to various learning skills. This research (e.g., Brown & Campione, 1977; Paris, Lipson, & Wixson, 1983; Ryan, 1981) indicates that successful use of knowledge in the regulation and control of learning processes is highly dependent on the ability of learners. The results of the present study are consistent with these findings, extending this work into the area of question answering strategies. The findings indicate that instructing students in the relationships between questions and sources of information improved their performance overall, particularly those of lower ability levels. This suggests a causal relationship (vis à vis Belmont & Butterfield, 1977) between the means taught and the goal of providing appropriate answer information.

The question arises as to what the QAR training specifically changed about the way in which students interacted with text. Certainly they did improve in their comprehension, but why? Three explanations are possible. Research in students' use of lookbacks during the process of reading and answering comprehension questions suggests that those students who are successful in answering questions, who are of high ability, use lookbacks more effectively (Garner & Reis, 1981). The finding that students of average and low ability levels in this study improved in their ability to answer questions that required text-based information suggests that this explanation is likely. A second explanation is that students used their background knowledge where appropriate, and their performance levels rose. This explanation is consistent only with the data from the higher ability students. The work of both Hansen and Pearson (1983) and Au (1979) suggests that lower ability students need specific instruction in how to use background knowledge, often needing the teacher to provide such knowledge before they read texts. Thus, although one would not expect that QAR training is sufficient for building knowledge, it may be sufficient to remind higher ability students to use the knowledge they already possess. A final explanation of what may have changed as a result of QAR instruction is that students' understanding of what it means to answer comprehension questions changed (i.e., the training did not change their comprehension processes, but changed the way they approached the task of answering questions assessing their comprehension). It is not possible from the data to determine the likelihood of this explanation. Certainly it is reasonable to conclude that making task demands of questions more apparent is likely to improve students' abilities to respond to questions. It is also likely that the answer to the question of how training affected students is a combination of the three explanations.

In summary, this study suggests that the knowledge of the existence and applicability of different sources of information is a part of students' prose comprehension processes. Theoretically, this adds to our understanding of the processes involved in question answering. In terms of implications for

instruction, this study suggests that basic to students' ability to apply information both from texts and from their knowledge base when answering comprehension questions is the importance of an awareness of the interplay among texts, background knowledge, and types of comprehension question.

REFERENCES

ANDERSON, R. C., & BIDDLE, W. B. (1975). On asking people questions about what they do when they are reading. In G. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9). New York: Academic Press.

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

AU, K. H., & IGNACIO, A. J. (1983, April). *Teacher questioning and metacognitive development in reading comprehension.* Paper presented at the annual meeting of the American Educational Research Association, Montreal.

BAKER, L., & BROWN, A. L. (1984). Metacognitive skills in reading. In P. D. Pearson (Ed.), *Handbook of reading research.* New York: Longman.

BARRETT, T. C. (1976). Taxonomy of reading comprehension. In R. Smith & T. C. Barrett (Eds.), *Teaching reading in the middle grades.* Reading, MA: Addison-Wesley.

BARTOLOME, P. (1969). Teacher's objects and questions in primary reading. *Reading Teacher, 23,* 27–33.

BELMONT, J. M., & BUTTERFIELD, E. C. (1977). The instructional approach to developmental cognitive research. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

BLOOM, B. B. (1956). *A taxonomy of educational objectives.* New York: McKay.

BOCK, R. D. (1975). *Multivariate statistical methods in behaviorial research.* New York: McGraw-Hill.

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

BROWN, A. L., CAMPIONE, J. C., & DAY, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher, 10,* 14–21.

DAY, J. D. (1980). *Teaching summarization skills: A comparison of training methods.* Unpublished doctoral dissertation, University of Illinois.

DURKIN, D. (1978–79). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14,* 481–533.

FRY, E. B. (1968). A readability formula that saves time. *Journal of Reading, 11,* 513–516, 575–578.

GARNER, R., & REIS, R. (1981). Monitoring and resolving comprehension obstacles: An investigation of spontaneous text lookbacks among upper-grade good and poor comprehenders. *Reading Research Quarterly, 16,* 569–582.

GORDON, C. J. (1979). *The effects of instruction in meta-comprehension and inferencing on children's comprehension abilities.* Unpublished doctoral dissertation, University of Minnesota.

GUSZAK, F. J. (1967). Teacher questioning and reading. *Reading Teacher, 21,* 227–234.

HANSEN, J., & PEARSON, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology, 75*(6), 821–829.

HARE, V. (1982). Beginning reading theory and comprehension questions in teachers' manuals. *Reading Teacher, 35,* 918–923.

LIPSON, M. Y. (1983). The influence of religious affiliation on children's memory for text information. *Reading Research Quarterly, 18*(4), 448–457.

PARIS, S. G., LIPSON, M. Y., & WIXSON, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8,* 293–316.

PEARSON, P. D. (1983). A critique of F. J. Guszak's study: Teacher questioning and reading. In L. M. Gentile, M. L. Kamil, & J. S. Blanchard (Eds.), *Reading research revisited* (pp. 271–281). Columbus, OH: Merrill.

PEARSON, P. D., HANSEN, J., & GORDON, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior, 11,* 201–209.

PEARSON, P. D., & JOHNSON, D. D. (1978). *Teaching reading comprehension.* New York: Holt, Rinehart, and Winston.

RAPHAEL, T. E., WINOGRAD, P., & PEARSON, P. D. (1980). Strategies children use when answering questions. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives on reading research and instruction.* Washington, DC: National Reading Conference.

RAPHAEL, T., & WONNACOTT, C. (in press). Heightening fourth grade students sensitivity to sources of information for answering comprehension questions. *Reading Research Quarterly.*

RICKARDS, J. P., & HATCHER, C. W. (1978). Interspersed meaningful learning questions as semantic cues for poor comprehenders. *Reading Research Quarterly, 13,* 539–553.

ROTHKOPF, E. Z. (1965). Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), *Learning and the educational process.* Chicago, IL: Rand-McNally.

ROTHKOPF, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3,* 241–249.

RYAN, E. B. (1981). Identifying and remediating failures in reading comprehension: Toward an instructional approach for poor comprehenders. In T. G. Waller & G. E. MacKinnon (Eds.), *Advances in reading research* (Vol. 2). New York: Academic Press.

SANDERS, N. (1966). *Classroom questions: What kinds?* New York: Harper & Row.

STERNBERG, R. J. (1981). Testing and cognitive psychology. *American Psychologist, 36,* 1181–1189.

TAYLOR, B. M. (1980). Children's memory for expository text after reading. *Reading Research Quarterly, 15,* 399–412.

TIERNEY, R. J. (1976). A comparison of Australian and American reading teachers. In J. T. Merritt (Ed.), *New horizons in reading.* Newark, DE: International Reading Association.

TIERNEY, R. J., BRIDGE, C., & CERA, M. J. (1978–79). The discourse processing operations of children. *Reading Research Quarterly, 14,* 539–573.

WHIMBEY, A. (1975). *Intelligence can be taught.* New York: Dutton.

WIXSON, K. K. (1984). Level of importance of postquestions and children's learning from text. *American Educational Research Journal, 21,* 419–434.

WONNACOTT, C. A. (1984). *Answering reading comprehension questions: The relationship of question-answering behaviors and metacognition with response quality.* Unpublished master's thesis, University of Utah.

WONNACOTT, C. A., & RAPHAEL, T. E. (1982, December). *Comprehension monitoring: An investigation of children's question-answering strategies.* Paper presented at the National Reading Conference, Clearwater, Florida.

AUTHORS

TAFFY E. RAPHAEL, Assistant Professor, Departments of Teacher Education and Educational Psychology, Michigan State University, 437 Erickson Hall, East Lansing, MI 48824. *Specialization:* Cognitive processes in reading.

P. DAVID PEARSON, Professor of Education, Center for the Study of Reading, University of Illinois, 51 Gerty Drive, Champaign, IL 61820. *Specialization:* Reading research.