DOCUMENT RESUME

ED 267 768 IR 012 136

AUTHOR Garhart, Casey; Hannafin, Michael

TITLE The Accuracy of Cognitive Monitoring during Computer-Based Instruction.

PUB DATE Jan 86

NOTE 22p.; Paper presented at the Annual Convention of the Association for Educational Communications and Technology (Las Vegas, NV, January 16-21, 1986). For entire proceedings, see IR 012 121.

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Academic Achievement; *Cognitive Processes; Cognitive Style; *Comprehension; *Computer Assisted Instruction; Higher Education; *Individualized Instruction; Learning Strategies; Predictor Variables; Regression (Statistics); Research Methodology; *Self Evaluation (Individuals); Tables (Data)

IDENTIFIERS AECT Research and Theory Division Meeting; Instructional Effectiveness; *Learner Control

ABSTRACT

This study was conducted to determine the accuracy of learners' comprehension monitoring during computer-based instruction and to assess the relationship between enroute monitoring and different levels of learning. Participants were 50 university undergraduate students enrolled in an introductory educational psychology class. All students received the same treatment--four sections of computer-based instruction designed to require learners to sort among numerous facts and concepts. Following each section, students were asked to rate their level of understanding of both the factual and the inferential materials. They then answered eight short-answer embedded questions--four factual and four inferential. Factual questions were divided into three levels of specificity and detail; inferential questions were divided into intraframe and interframes inferences. A posttest of 32 items was also administered; half of the questions were repeated from the lesson and half were new items. Regression procedures were conducted to determine the predictive value of enroute self-assessments for corresponding factual and inferential learning. Enroute ratings were used to predict student performance on embedded and posttest questions, as well as to examine the relationship among ratings for facts or inferences. Little or no correlation was found between the ratings of understanding and subsequent performance on either the embedded questions or the posttest, suggesting that learners may not be good judges of their enroute comprehension. A list of references and nine data tables complete the study. (JB)

The Accuracy of Cognitive Monitoring During Computer-Based Instruction

Casey Garhart and Michael Hannafin

Center for Research and Development in Education Computing

The Pennsylvania State University

176 Chambers

University Park, PA 16802

Running Head: COGNITIVE MONITORING

Submitted September 25, 1985

ABSTRACT

In order to individualize instruction, computer-based lessons often allow the learner to determine which material will be covered, and the sequence to be followed. Unfortunately, subjects \ ontrol their instructional decisions often perform worse than subjects .der program control. This study examined the relationship between self-assessed understanding of lesson content and performance on factual and inferential test questions among 50 college undergraduates. Analysis showed little or no correlation between ratings of understanding and subsequent preformance on both embedded questions and a posttest. This apparent inability to accurately assess understanding may help to explain why subjects who control their own instructional decisions tend to terminate instruction prematurely

THE ACCURACY OF COGNITIVE MONITORING DURING COMPUTER-BASED INSTRUCTION

Individuals process information in different ways. The strategies preferred by one learner are likely to differ from those preferred by another. Ideally, every lesson should be individually tailored to suit the needs and abilities of each learner "so as to develop, compensate for, or capitalize upon student characteristics for the optimization of subject-matter learning" (Messick, 1984, p. 69). One way to accomplish such individual tailoring is by transferring control of the lesson's structure and sequence to the learner. The greater the learner control, the more individualized instruction should become.

Unfortunately, research in this area has failed to demonstrate that learner control consistently improves learning. Instead, studies indicate that imposed control increases learning significantly over programs in which the learner controls the instructional scope and sequence (See, for example, Atkinson, 1972; Park & Tennyson, 1983; Ross & Rakow, 1981; Tennyson, Tennyson & Rothen, 1980; Tennyson & Buttrey, 1980; Tennyson, Christensen & Park, 1984). Tennyson (1980, p. 505) stated that "instructional research dealing with variables of learner control has failed to demonstrate that students can make and carry out decisions of content element selection and personal learning assessment."

A basic problem noted by locus on instructional control researchers has been that subjects who controlled their instruction frequently terminated instruction prematurely. However, the cause of this phenomenon is not clear. If subjects who control instruction direct as much effort to learn as those under program control, then other factors must affect instructional control decisions.

One possible explanation is that many learners do not assess accurately their understanding of lesson information. Less skilled learners, and perhaps learners who encounter new subject matter, tend not to detect their failure to understand new material (August, Flavell & Clift, 1984; Baker, 1979; Flavell, 1979; Garner, 1981; Garner & Anderson, 1982; Grabe & Mann, 1984; Markman, 1977; Robinson & Robinson, 1984; Whimbey, 1976).

If instructional control decisions are made based upon accurate student comprehension monitoring, successful outcomes should be expected; if based upon monitoring that is ineffective, however, the outcomes of learner-directed lessons are likely to be poor. The accuracy of student comprehension monitoring, therefore, is a critical variable in learner-directed computer-based instruction.

Deese (1969) suggested that understanding is an introspective process that each person is capable of recognizing. This is supported by Hart's (1967) work demonstrating that perceptions of comprehension were very accurate when compared with a subsequent recognition posttest. Several studies, however, contradict Deese and Hart's hypotheses, indicating that learners are not always good judges of their understanding (August, Flavell & Clift, 1984; Baker, 1979; Bransford & Nitsch, 1978; Brown, Campione & Barclay, 1979; Garner & Anderson, 1982; Goetzfried & Hannafin, 1985; Markman, 1977; Robinson & Robinson, 1984; Whimbey, 1976). In these studies subjects either thought they understood or ignored their lack of comprehension during instruction.

The types and levels of learning are also likely to influence both self-assessment and en-route comprehension. Numerous studies have shown that main ideas are better remembered than details (Britton, Simpson, Meyer, Holdredge & Curry, 1979; Johnson, 1970; Kintsch, 1974; Kintsch & Keenan, 1973; Meyer, 1975; Meyer & Rice, 1981; Walker & Meyer, 1980). It may be the understanding of these main ideas upon which subjects base self assessments. Meyer and others developed a system for identifying the level of propositions in text, and concluded that information high in the structure (main, or superordinate ideas) is more likely to be integrated into the learner's cognitive schema than information low in the structure (details, or subordinate ideas).

In this study the accuracy of student's comprehension monitoring during computer-based instruction, and the relationship between enroute monitoring and different levels of learning, were studied.

METHODS

Subjects

The subjects were 50 university undergraduates enrolled in an introductory educational psychology class at a large public university. Subjects participated in the study on a voluntary basis and were awarded extra credit for their participation.

Materials

The lesson concerned the discovery of a fictional ore (berkelium oxide) on an imaginary South Seas island. Although the details of the material were fictitious in order to avoid the influences of prior knowledge, the content was designed to avoid logical conflicts with concepts pertaining to history, economics, anthropology, and mineral science.

The information was divided into four sections and presented in the form of computer-based instruction. The four sections were *The History of Jexium Island*, *The Discovery of Berkelium Oxide*, *The Mining of Berkelium Oxide*, and *The Market for Berkelium Oxide*. The average number of frames per section was 11 with a maximum of 13 and a minimum of 10. Each section also contained two graphic drawings which were used to maintain motivation, but did not relate to items later tested. The material was designed to include a high information density within individual frames and sections in order to require students to sort among numerous facts and concepts.

Following each section, students were asked to rate their level of understanding of both the factual and inferential material on scales from one (not at all) to five (very well). They then answered eight short-answer embedded questions pertaining to the material. The eight embedded questions consisted of four factual questions and four inferential questions. Questions following each section covered only information presented in that section and were not cumulative.

Factual Lesson Content

Factual questions were those which required the recall of information which had been stated explicitly in the text. The questions were divided into three levels, based on elements common to the parsing hierarchies of both Meyer and Kintsch. The system used was considered more appropriate for the individual frames of computer-based instruction. Factual questions were divided into three Level 1 questions, eight Level 2 questions, and five Level 3 questions.

Level 1. Level 1 questions were the most general and could be answered with the main idea of a computer frame. An example of a level 1 question is: "What was the ore Groningen discovered?"

Level 2. Level 2 questions were more specific and required information which supported the main ideas. An example of a level 2 question is: "Why was the discovery of the ore important to laboratory scientists?"

Level 3. Finally, level 3 questions were the most specific and covered details which were less significant to the main ideas of the story. An example of a level 3 question is: "In what decade did Groningen go to Jexium Island?"

Inferential Lesson Content

Inferential questions required students to evaluate two or more related pieces of information in order to form a conclusion. The questions were divided into intraframe and interframe inferences. Intraframe and interframe inferential questions were distributed evenly across sections.

Intraframe. Intraframe inferences could be answered based upon information presented in a single frame. An example of an intraframe question is: "What important event occured on Jexium Island in 1945?"

Interframe. Interframe inferences required that information from two or more frames be evaluated simultaneously in order to form an appropriate inference. An example of an interframe question is: "Who built the towns around the mine site?"

Lesson Posttest

There were a total of 32 items on the posttest: half were repeated from the lesson and half were new items. The information tested using the new questions was evenly divided across the four sections and contained equal numbers of factual and inferential questions. Sixteen of the questions were factual and 16 were inferential, yielding four learning measures: Repeated Facts, New Facts, Repeated Inferences, New Inferences. One-half of the inference questions were interframe and the others were intraframe questions. Three of the factual questions were Level 1, seven were Level 2, and six were Level 3. The items were presented in random order. All test items were short-answer type questions.

Dependent Measures

Several dependent measures were collected. Dependent measures related both to students' ratings of their understanding and to their subsequent performance on both embedded and posttest questions were collected.

Separate measures were obtained for fact and inference ratings for each of the sections of the lesson. In addition, the number of correct responses to embedded factual and inferential questions was computed, as well as an aggregated fact and inference score for the lesson.

Student performance on the posttest was organized in two ways. First, correct answers were tallied to produce a fact and an inference scale. Next, factual items were classified as Level 1, Level 2, or Level 3, and inference items were classified as Interframe or Intraframe.

Design and Data Analysis

The design was a complete repeated measures design, where all participants received exactly the same instructional treatment. All dependent measures were gathered on all students. Regression procedures were conducted to determine the predictive value of en-route self-assessments for corresponding factual and inferential learning both for the embedded and the posttest questions.

Enroute ratings for factual understanding were used to predict student performance on embedded factual questions while enroute ratings for inferences were used to predict student performance on embedded inference questions. Enroute ratings were also used to predict posttest performance for each corresponding scale. In addition, enroute ratings were intercorrelated in order to examine the relationship among ratings for facts or inferences.

Procedures

All students received the same treatment. Students reported to a microcomputer lab during one of eight periods reserved specifically for the study. They were told the study was designed to investigate how well people understand material presented via computer, what they understand, and how well they can evaluate their own understanding. They were also told that the lesson consisted of four parts with approximately 10 frames of information contained in each section and that they would be asked short answer questions over the material after each section. This information was given orally at the beginning of each session and also repeated at the beginning of the lesson. Since student input was recorded during the computer program, students were told not to be concerned with correct spelling, but to confine answers to a single line. No time limit was imposed for responding to the questions.

Students then completed the lesson. During each section they viewed the lesson and answered the eight embedded questions. Students did not receive any knowledge of their results on these questions. Following all four sections, the 32 item posttest was completed.

RESULTS AND DISCUSSION

Factual Ratings with Performance

Factual ratings correlated with embedded scores. Table 1 contains correlations between students' ratings of their factual understanding and their scores on the factual questions following each section of the lesson. Summed across sections, the total factual ratings and performance on factual embedded questions were correlated at .30 ($p < .05$). However, most of this correlation could be accounted for by performance on the fourth section. During the first three sections, the correlations between ratings and performance were not significant. Only during the fourth section were ratings and scores on factual questions significantly correlated at .30 ($p < .05$).

Insert Table 1 about here.

This could indicate that students improved in their ability to rate understanding as a result of the embedded questions. However, although the most accurate predictions were made in the fourth section, the effect was not progressive during the first three sections.

Intercorrelations among factual ratings. Table 2 contains correlations between ratings of factual understanding in each of the four sections. Factual ratings were intercorrelated fairly well (minimum $p < .01$), indicating that students use an internally consistent system to rate their understanding of factual information, but that the system was not as highly related to their actual knowledge of the information tested.

Insert Table 2 about here.

Correlations were generally highest as students progressed chronologically. Self ratings for Section 1 correlated more highly with ratings for Section 2 (.55) than with ratings for later sections. Ratings for Section 3 correlated highest with Section 4 (.59). This might indicate that adjustments in assessing understanding were made gradually as students progressed through the lesson, modifying their criteria for judgement based on experience obtained during the lesson.

Factual ratings correlated with posttest scores. Table 3 contains correlations between enroute ratings of factual understanding and posttest scores on questions from each of the four sections. Although students performed approximately as well on the posttest questions as on the embedded questions ($r = .72$, $p < .0001$), correlations between ratings and scores were generally lower than during the lesson and were even negatively correlated in some instances.

Insert Table 3 about here.

Factual ratings correlated with levels of factual questions. Table 4 contains correlations between ratings and performance on the three levels of fact questions. Because levels were not evenly distributed among sections and there were only a few examples of each level within a section, total scores are used.

Insert Table 4 about here.

The only significant correlations were found between Section 1 ratings and performance during the lesson for each level. These correlations are also reflected in the overall correlation between Section 1 ratings and total score shown in Table 1. The general consistency across levels, especially as seen in the total ratings correlated with levels, may indicate a consistency in ratings across levels of information. The slightly lower correlations for Level 1 questions throughout may be accounted for by the fact that there were simply fewer Level 1 questions. Although correlations with Level 3 questions on the posttest were slightly higher than for the other two levels, they were generally not significant.

Inference Ratings with Performance

Inference ratings correlated with embedded scores. Table 5 contains correlations between ratings of inferential understanding and scores on the embedded inference questions. Summed across sections, the total inferential ratings and performance on inferential embedded questions were correlated at .38 ($p < .01$). As with the factual correlations, however (see Table 1), most of this could be accounted for by performance on later sections. The general trend for correlations between ratings and performance on individual sections was similar to the trend found for factual questions. Initially, correlations were low, but in section three ratings and performance were significantly correlated at .37 ($p < .01$). However, correlations declined in Section 4.

Insert Table 5 about here.

Intercorrelations among inference ratings. Table 6 contains correlations between students' ratings of inferential understanding in each of the four sections. In all but one instance (Section 3 with Section 1), ratings were significantly intercorrelated. The correlation among inference ratings was greater than for performance on either embedded or posttest inference questions. Again, this suggests that students use an internally consistent system to rate their understanding but that this system is not related favorably to tested knowledge.

Insert Table 6 about here.

Inference rating intercorrelations exhibited the same general trends as factual rating intercorrelations. This trend may indicate that students modify their criteria for self-assessment based on lesson experiences. However, it may required several lesson sections to form a reliable system for judging inference.

Inference ratings correlated with posttest scores. Table 7 contains correlations between enroute ratings of inference understanding and posttest scores based on information from each of the four sections. As with factual questions, students performed approximately as well on the inference posttest questions as they had on the embedded questions ($r = .71$, $p < .0001$). The trends between ratings and performance varied, however, from embedded questions to the posttest

Insert Table 7 about here.

Correlations were strongest between ratings and performance on Sections 1 and 4, with a moderate correlation for Section 3 and virtually no correlation for Section 2. The global, seemingly random, relationships between ratings and performance were typified by the correlations obtained between non-aligned section reatings and scores. Significant correlations were found between Section 4 ratings and Section 1 performance ($p < .01$) and between Section 3 ratings and Section 4 performance ($p < .05$).

Inference ratings correlated with type of inference question. Table 8 contains correlations between ratings and performance on the two types of inference questions (Interframe and Intraframe). Because there were only four examples of each question type per section, total scores were used. Significant differences for Sections 3 and 4 account for most of the correlation between ratings and scores for both embedded and posttest questions.

Insert Table 8 about here.

Student performance on en-route inter and intra-frame inferences was very similar (61% and 63.75%). It is therefore unlikely that differences in correlations could be attributed to differences in performance. Students appear to judge their inference comprehension more or less singularly, and do not seem affected particularly by either within- or interframe influences.

Fact and Inference Ratings by Section

Table 9 contains intercorrelations between ratings of factual understanding and inferential understanding for each section. Ratings within sections produced the highest correlations of the study (.48, .45, .67, and .78 respectively) and were all significant at the .001 level, despite the fact that performance on fact and inference questions within sections did not correlate highly. This supports the assumption that although students were using some system for judging their level of understanding, the system was not highly related to their actual knowledge of the kinds of information tested. The system also did not differentiate effectively between factual and inferential learning. Rather, both fact and inference ratings appeared related to some global criteria on which students based their assessments.

Insert Table 9 about here.

GENERAL DISCUSSION

The results of this study suggest that learners are not good judges of their en-route comprehension. Little or no correlation between ratings of understanding and subsequent performance for both factual and inferential material were found for either embedded questions or the posttest.

One possible explanation for these results may be related to the subjective nature of "understanding." Students may assess their understanding according to criteria different from one another as well as from the experimenter (Baker, 1979; Garner & Anderson, 1982). Though internally consistent both within and between ratings for fact and inference, the ratings are not related well to any of the scales employed in this study. In addition, the more or less random correlations with the different levels of factual and inferential learning suggest that the student ratings were not based on the types of learning addressed in this study. Ratings appear to be based more on the undifferentiated, global perceptions of students as to their understanding. This presents a potential problem in practice, where the specific intended lesson information may not be the basis for making learner-based instructional control decisions. Based on the findings of the present study, it is simply unclear as to what information is used by students to estimate comprehension.

Some degree of acclimation to the lesson content and procedures was presumed necessary before self-assessments could be considered valid. After several sections and attempts to answer questions, student ratings of understanding should be more accurate, and successive ratings more highly correlated with actual performance. Although this pattern was not demonstrated completely, the data indicated some trends in this direction. Correlations between ratings and performance for both factual and inferential questions were significant mainly in later sections. This trend might have been more pronounced if understanding of factual and inferential questions were more consistent across students. This might be accomplished by clarifying the rating task more through additional initial instruction, including examples of factual and inferential questions, specifying explicitly which questions were factual or inferential, asking students for ratings of specific facts or inferences, or providing response feedback.

Individuals may also evaluate understanding at levels other than those selected in this study. For example, learners may assess understanding correctly at low levels but fail to demonstrate understanding at a higher level. Low level assessment of understanding should be reflected in higher correlations of ratings with factual questions than with inferential questions. It was expected that ratings would correlate more highly with main ideas (gist level) than with lower level facts, indicating that students based their assessment on knowledge of general ideas. The data, however, did not reflect any significant relationships between assessment and level of factual information.

An assessment of understanding based on a high level of assimilation of the material should have caused ratings to be more highly correlated with inferential questions than with factual questions. In effect, one might predict that ratings of inference would be the best predictors of student performance. Again, however, this was not demonstrated. None of the self-assessments were found to be uniformly more accurate than others in forecasting student performance. Unfortunately, ratings were not highly correlated with performance on any of the scales. Instead, there were only correlations between the ratings themselves. Scores on both the embedded questions and the posttest indicated that students remembered high level questions best, but they did not rate their level of understanding based on an accurate assessment of this knowledge.

Finally, since understanding involves the integration of new information with prior knowledge, studies which are relatively short and cover only a small amount of new information may not give learners sufficient time to develop new, or to adapt existing, schema effectively. If an inability to correctly assess understanding is related to the lack of an integrated cognitive schema, assessment and performance should improve in later sections.

Cognitive monitoring can be particularly difficult to study. The process can only be inferred from observed outcomes and from the introspective reports of subjects. [illegible]ts who lack experience with the process of introspection may be u[illegible] of how to attend or what the focus of the attention should be. The resulting reports may reflect processes not anticipated by the experimenter. The findings of this study may be related to such problems.

The popularity of learner controlled computer-based instruction accentuates the importance of further cognitive monitoring research. The tendencies reported for premature withdrawal from CBI lessons may be associated with basic misperceptions of learning. The lack of stronger correlations between self-assessed understanding and actual performance indicates the need for further research in this area.

REFERENCES

Atkinson, R.C. (1972) Optimizing the learning of a second-language vocabulary. Journal of Experimental Psychology, 96, 124-129.

August, D.L., Flavell, J.H., & Clift, R. (1984) Comparison of comprehension monitoring of skilled and less skilled readers. Reading Research Quarterly, 20, 39-53.

Bransford, J.D. & Nitsch, K.E. (1978) Coming to understand things we could not previously understand. In J.F. Kavanagh and W. Strange (Eds)., Speech and language in the laboratory, school and clinic. Cambridge, MA: MIT Press.

Britton, B.K., Simpson, R., Meyer, B.J.F., Holdredge, T.S., & Curry, C. (1979) Effects of the organization of text on memory: Tests of two implications of a selective attention hypothesis. Journal of Experimental Psychology: Human Learning and Memory, 5, 496-506.

Brown, A.L., Campione, J.C. & Barclay, C.R. (1979) Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall. Child Development, 50, 501-512.

Deese, J. (1969) Behavior and fact. American Psychologist, 24, 515-522.

Flavell, J.H. (1979) Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist, 34, 906-911.

Garner, R. (1981) Monitoring of passage inconsistency among poor comprehenders: A preliminary test of the "piecemeal processing" explanation. Journal of Educational Research, 74, 159-162.

Garner, R. & Anderson, J. (1982) Monitoring of understanding research: Inquiry directions, methodological dilemmas. Journal of Experimental Education, 50, 70-76.

Goetzfried, L. & Hannafin, M.J. (1985) The effect of the locus of CAI control strategies on the learning of mathematics rules. American Educational Research Journal, 22, 273-278.

Hart, J.T. (1967) Memory and memory monitoring processes. Journal of Verbal Learning and Verbal Behavior, 6, 685-691.

Johnson, R.E. (1970) Recall of prose as a function of the structural importance of linguistic units. Journal of Verbal Learning and Verbal Behavior, 9. 12-20.

Kintsch, W. (1974) The representation of meaning in memory. Potomac, MD: Lawrence Erlbaum Associates.

Kintsch, W. & Keenan, J.M. (1973) Reading rate as a function of number of propositions in the base structure of sentences. Cognitive Psychology, 6, 257-274.

Markman, E.M. (1977) Realising that you don't understand: A preliminary investigation. Child Development, 48, 986-992.

Messick, S. (1984) The nature of cognitive styles: Problems and promise in educational practice. Educational Psychologist, 19, 59-74.

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

Meyer, B.J.F. & Rice, G.E. (1981) Information recalled from prose by young, middle, and old adult readers. Experimental Aging Research, 7, 253-268.

Park, O. & Tennyson, R.D. (1983) Computer-based instructional systems for adaptive education: A review. Contemporary Education Review, 2, 121-135.

Robinson, E.J. & Robinson, W.P. (1984) Realizing you don't understand: A further study. Journal of Child Psychology and Psychiatry and Allied Disciplines, 25, 621-627.

Ross, S.M & Rakow, E.A. (1981) Learner control vs. program control as adaptive strategies for selection of instructional support on math rules. Journal of Educational Psychology, 73, 745-753.

Tennyson, C.L., Tennyson, R.D., & Rothen, W. (1980) Content structure and management strategy as design variables in concept acquisition. Journal of Educational Psychology, 72, 499-505.

Tennyson, R.D. (1980) Instructional control strategies and content structure as design variables in concept acquisition using computer-based instruction. Journal of Educational Psychology, 72, 525-532.

Tennyson, R.D. & Buttrey, (1980) Advisement and management strategies as design variables in computer-assisted instruction. Eductional Communication and Technology Journal, 28, 169-176.

Tennyson, R.D., Christensen, D.L. & Park, S.I. (1984) The Minnesota Adaptive Instructional System: An intelligent CBI system. Journal of Computer-Based Instruction, 11, 2-13.

Walker, C.H. & Meyer, B.J.F. (1980) Integrating different types of information in text. Journal of Verbal Learning and Verbal Behavior, 19, 263-275.

Whimbey, T. (1976) Intelligence can be taught. New York: Bantam.

Table 1
Ratings of factual understanding correlated with scores on factual questions for each section

Section Scores

Fact Ratings	Sect 1	Sect 2	Sect 3	Sect 4	Total
Sect 1	.25	.22	.20	.17	.38**
Sect 2	.11	-.05	.43**	.14	.23
Sect 3	-.02	-.01	.10	.14	.07
Sect 4	.02	.11	.33*	.30*	.26

Note. Overall fact ratings correlated with fact total at .30, $p < .05$.

*$p < .05$ **$p < .01$

Table 2
Intercorrelations among factual ratings across sections

Fact Ratings

Fact Ratings	Sect 1	Sect 2	Sect 3	Sect 4
Sect 1	X	.55	.35	.44
Sect 2	(.0001)	X	.46	.47
Sect 3	(.01)	(.001)	X	.59
Sect 4	(.001)	(.001)	(.0001)	X

Table 3
Ratings of factual understanding correlated with scores on factual posttest questions by section

Fact Ratings	Posttest Scores: Sect 1	Sect 2	Sect 3	Sect 4
Sect 1	.16	-.03	.16	.33*
Sect 2	.12	-.16	.20	.21
Sect 3	.03	-.16	.17	.01
Sect 4	.04	-.03	17	.17

* $p < .05$

Table 4
Ratings of factual understanding correlated with scores on factual questions by level

Fact Ratings	Levels of Fact Questions: Embedded Totals			Posttest		
	1	2	3	1	2	3
Sect 1	.28*	.33*	.29*	.12	.10	.29*
Sect 2	.05	.23	.17	02	.07	.22
Sect 3	-.02	-.01	.11	12	-.10	.11
Sect 4	.07	.24	.24	13	.00	.19
Total	.12	.25	.26	.13	.02	.26

* $p < .05$

Table 5
Ratings of inference understanding correlated with scores on embedded inference questions by section

Inference Ratings	Section Scores: Sect 1	Sect 2	Sect 3	Sect 4	Total
Sect 1	.10	-.14	.25	.17	.17
Sect 2	.06	.12	.12	-.08	.10
Sect 3	.23	.09	.37**	.28	.49**
Sect 4	.27	-.08	.30*	.24	.31*

Note. Overall inference ratings correlated with inference total at .38, $p < .01$.

* $p < .05$ ** $p < .01$

Table 6
Intercorrelations among inference ratings

Inference Ratings	Inference Ratings: Sect 1	Sect 2	Sect 3	Sect 4
Sect 1	X	.29	.10	.35
Sect 2	(.05)	X	.45	.30
Sect 3	NSD	(.001)	X	.49
Sect 4	(.01)	(.05)	(.0001)	X

Table 7
Ratings of inferential understanding correlated with scores on inferential posttest questions by section

		Posttest Scores			
		Sect 1	Sect 2	Sect 3	Sect 4
Inference Ratings	Sect 1	.30*	.25	-.21	.03
	Sect 2	.05	.07	.08	.06
	Sect 3	.23	.13	.24	.36*
	Sect 4	.41**	.02	.21	.34*

* $p < .05$ ** $p < .01$

Table 8
Ratings of inference understanding correlated with scores on inference questions by type

		Types of Inference Questions			
		Embedded Totals		Posttest	
		Inter	Intrs	Inter	Intrs
Inference Ratings	Sect 1	.24	.01	.04	.16
	Sect 2	.11	00	.06	.11
	Sect 3	.48**	37*	[illegible]	.36**
	Sect 4	26*	22	.18	.44**
	Total	.39**	.22	.20	.39**

* $p < .05$ ** $p < .01$

Table 9

Ratings of factual understanding correlated with ratings of inference understanding

Fact Ratings	Inference Ratings: Sect 1	Sect 2	Sect 3	Sect 4
Sect 1	.48**	.07	.30*	.42**
Sect 2	.21	.45**	.34*	.36**
Sect 3	.02	.29*	.67**	.41**
Sect 4	.40**	.29*	.49**	.78**

* $p < .05$ ** $p < .01$