

## Question Asking During Tutoring

Arthur C. Graesser and Natalie K. Person  
*Memphis State University*

*Whereas it is well documented that student question asking is infrequent in classroom environments, there is little research on questioning processes during tutoring. The present study investigated the questions asked in tutoring sessions on research methods (college students) and algebra (7th graders). Student questions were approximately 240 times as frequent in tutoring settings as classroom settings, whereas tutor questions were only slightly more frequent than teacher questions. Questions were classified by (a) degree of specification, (b) content, and (c) question-generation mechanism to analyze their quality. Student achievement was positively correlated with the quality of student questions after students had some experience with tutoring, but the frequency of questions was not correlated with achievement. Students partially self-regulated their learning by identifying knowledge deficits and asking questions to repair them, but they need training to improve these skills. We identified some ways that tutors and teachers might improve their question-asking skills.*

---

ARTHUR C. GRAESSER is a Professor of Psychology and Mathematical Sciences in the Department of Psychology at Memphis State University, Memphis, TN 38152. His specializations are discourse processes and questioning mechanisms.

NATALIE K. PERSON is a graduate student in the Department of Psychology at Memphis State University, Memphis, TN 38152. Her specializations are tutoring, teaching strategies, and questioning mechanisms.

---

Question asking has had a controversial status in education. At one extreme, some researchers believe that question asking (and answering) are very central components in theories of learning, cognition, and education. Ideal students presumably are capable of actively self-regulating their learning by being sensitive to their own knowledge deficits and by seeking information that repairs such deficits. Researchers in education and developmental psychology have frequently advocated educational settings that engage students in active learning and problem solving, or that directly train self-regulatory learning strategies in students (Bransford, Arbitman-Smith, Stein, & Vye, 1985; Collins, 1985, 1988; Palincsar & Brown, 1984; Papert, 1980; Piaget, 1952; Pressley & Levin, 1983; Pressley, Goodchild, Fleet, Zajchowski, & Evans, 1989; Zimmerman, 1989). According to a number of models in cognitive science, question generation is a fundamental component in cognitive processes that operate at deep conceptual levels, such as the comprehension of text and social action (Collins, Brown, & Larkin, 1980; Hilton, 1990; Olson, Duffy, & Mack, 1985), the learning of complex material (Collins, 1985; Miyake & Norman, 1979; Palincsar & Brown, 1984; Schank, 1986), problem solving (Reisbeck, 1988), and creativity (Sternberg, 1987). There also is empirical evidence that improvements in the comprehension, learning, and memory of technical material can be achieved by training students to ask good questions (Davey & McBride, 1986; Gavelek & Raphael, 1985; King, 1989, 1990; Palincsar & Brown, 1984; Singer & Donlan, 1982).

At the other extreme, however, there are researchers who have reconciled themselves to the fact that student questions do not play a central role in the normal process of learning (Dillon, 1988). It is well documented that student questions in the classroom are very infrequent and unsophisticated. The estimated frequencies of student questions per hour have ranged from 1.3 to 4.0, with a median of 3.0, in available classroom studies (Buseri, 1988; Dillon, 1988; Fenclova, 1978; Good, Slavings, Harel, & Emerson, 1987; Nickel & Fenner, 1974; Susskind, 1969). These classroom studies have spanned several countries (including Germany, the United States, Nigeria, and Czechoslovakia), so the low frequency of student questions appears to be a universal phenomenon. Given that an average class has approximately 26.7 students (see Dillon, 1988), the frequency of questions generated by a particular student within an hour is only .11 question (i.e.,  $3 \div 26.7$ ). In contrast to the student questions, the frequency of teacher questions is quite high, ranging from 30 to 120 questions per hour, with a mean of 69 questions. Therefore, 96% of the questions in a classroom environment are teacher questions. In addition to being infrequent, student questions are also unsophisticated (Dillon, 1988; Flammer, 1981; Kerry, 1987). That is, they are normally shallow, short-answer questions that address the content and interpretation of explicit material; they are rarely high-level questions that involve inferences, multistep reasoning, the application of an idea to a new domain of knowledge, the synthesis of a new idea from multiple information sources, or the critical evaluation of a claim.

The low frequency and sophistication of student questions can be attributed to barriers at three different levels. One barrier is that students have difficulty identifying their own knowledge deficits (Baker, 1979; Glenberg, Wilkinson,

& Epstein, 1982; Graesser & McMahan, 1993; Markman, 1979; Pressley, Ghatala, Woloshyn, & Pirie, 1990), unless they have high amounts of domain knowledge (Brown, Bransford, Ferrara, & Campione, 1983; Chi, Bassok, Lewis, Reiman, & Glaser, 1989; Miyake & Norman, 1979). Students have difficulty detecting contradictory information, in identifying missing data that are necessary for a solution, and in discriminating superfluous from necessary information. A second barrier to question asking involves social editing (Graesser, McMahan, & Johnson, in press; van der Meij, 1987, 1988). The student reveals ignorance and loses status when a bad question is asked. There are social barriers even when a good question is asked, such as interrupting the teacher and changing the topic of conversation. A third barrier lies in a deficit in acquiring good question-asking skills. Most teachers are not excellent role models for asking good questions. A small percentage of teacher questions (4%) are high-level questions; most of their questions are short-answer questions that grill the students on explicit material (Kerry, 1987; Dillon, 1988). Few teachers adopt sophisticated Socratic methods in which the teacher constructs sequences of thought-provoking questions that expose the student's misconceptions and contradictions (Collins, 1985; Stevens, Collins, & Goldin, 1982). Given that several students are in a classroom, it would be impractical for a teacher to diagnose and repair the knowledge deficits of several particular students. For whatever reason, the low frequency and quality of student questions in a classroom constitute a challenge for educational theories and philosophies that emphasize active self-regulated learning.

It is conceivable that one-to-one tutoring environments might remove many of the barriers that prevent students from asking questions in a classroom. First, tutoring sessions are allegedly tailored to the knowledge deficits of a particular student. The tutor therefore has a sustained opportunity to identify and query such deficits. Second, many of the social barriers to asking questions are removed in a one-to-one tutoring session because there is a dialogue between only two individuals (i.e., tutor and student). The student may be embarrassed in front of his or her peers when a question reflects ignorance in a classroom setting, whereas pressure from peers is minimized in one-to-one tutoring. It is appropriate for the student to interrupt the tutor with questions and to change the topic of conversation in a student-centered exchange. Third, students might become exposed to better questions in a tutoring environment because the tutors have the opportunity to concentrate on deeper levels of understanding and reasoning. The present investigation of question asking was inspired by the possibility that tutoring environments hold some promise in facilitating question asking and active learning.

Surprisingly, few studies have investigated the process and educational benefits of tutoring. It is reasonably well documented that learning of mathematics and readings skills is better in one-to-one tutoring sessions than in classroom settings (Bloom, 1984; Cohen, Kulik, & Kulik, 1982). According to Cohen et al.'s (1982) meta-analysis of 52 tutoring studies, tutoring raised the performance of students by .4 standard deviation units compared to a conventional classroom control. It is informative to note that the impact of tutoring

on learning was not significantly related to the amount of tutoring training that the tutors had received or to age differences between tutor and student, but it was higher for structured than unstructured tutoring programs (see also Fitz-Gibbon, 1977). Tutoring is sometimes successful when tutors are peers of students rather than topic experts (Cohen et al., 1982; Rogoff, 1990). Perhaps tutors need to be exceptionally well trained in the tutoring process and in the domain knowledge before such expertise yields high payoffs in student learning. It should be noted, however, that most tutors in a school system consist of students, paraprofessionals, and adult volunteers rather than highly skilled tutors.

It is presently unclear why there is such an advantage of tutoring over classroom settings because of the lack of research on the process of tutoring (see Graesser, 1993, in press; McArthur, Stasz, & Zmuidzinas, 1990; Putnam, 1987). However, several hypotheses would account for the advantage. According to an *active inquiry* hypothesis, students perhaps ask more questions in tutoring sessions and thereby correct their own idiosyncratic knowledge deficits: Better learning would result from the inquiry and monitoring of knowledge deficits. According to an *explanation* hypothesis, tutoring may make explicit some important patterns of reasoning and problem solving that the classroom setting cannot readily furnish because of time and resource limitations. Explanations are provided by the tutor, by the student, or by a collaboration between tutor and student. Learning is facilitated to the extent that there are explanations of facts and events conveyed in the material to be learned (Anderson, Conrad, & Corbett, 1989; Chi et al., 1989; Pressley, Symons, McDaniel, Snyder, & Turture, 1988). Much of the reasoning and problem solving is exposed when deep-level questions are asked and answered (i.e., why, why not, how, what if). Perhaps explanatory reasoning and these deep-level questions are more prevalent in tutoring sessions and can account for the benefits from tutoring. According to a *questioning-skill* hypothesis, tutors may ask better questions in a tutoring setting than in a classroom, and these questioning skills may be learned by the student during the course of tutoring. As discussed earlier, there is ample evidence that learning improves to the extent that students learn how to ask good questions. According to a *diagnosis and remediation* hypothesis, tutoring provides more opportunities than classroom interactions for individual students to reveal their faulty thinking or misconceptions, and for the tutor to tailor subsequent instruction to the idiosyncratic needs of a particular student (van Lehn, 1990). Some potential explanations of the advantages of tutoring are unrelated to questioning *per se*. For example, perhaps students are merely more engaged with the material, for longer time spans, in tutoring environments. Although we acknowledge the potential importance of these other variables, the present study focuses on the role of questions in the tutoring process.

Given the lack of research on the tutoring process, the present study investigated the frequency and quality of questions that tutors and students ask during tutoring sessions. Transcripts of tutoring sessions were collected from college students in a scientific research methods class and from seventh graders having difficulties in algebra. Our analyses of questions focused on three specific objectives. The first objective was to document the frequency of questions dur-

ing tutoring and to compare these frequencies to those in classroom settings. The second objective was to analyze the qualitative characteristics of questions in the corpus, as will be discussed shortly. The third objective was to examine the extent to which the students' achievement in the research methods course was correlated with the frequency and qualitative characteristics of the student questions. It should be noted that it was well beyond the scope of this study to assess whether particular characteristics of the questioning process have a *causal* relationship with learning outcomes.

### Theoretical Schemes for Analyzing Questions

Schemes for analyzing the qualitative characteristics of questions have been proposed by researchers in education (Dillon, 1984, 1988; Flammer, 1981; van der Meij, 1987), psychology (Clark & Schaefer, 1989; Graesser, Langston, & Baggett, in press; Graesser, Person, & Huber, 1992), computational linguistics (Dalhgren, 1988; Gordon & Lakoff, 1972), and artificial intelligence (Allen, 1983; Kaplan, 1983; Lehnert, 1978; Schank, 1986; Webber, 1988). The present study adopted an analytical scheme developed by Graesser et al. (1992). The Graesser-Person-Huber (GPH) scheme was influenced both by the theoretical work in these fields and by preliminary empirical analyses of tutoring data. According to the GPH scheme, a question is analyzed on three dimensions: the content of the information requested, the psychological mechanism that generates a question (a dimension that considers the context of the question), and degree of specification. Aside from these theoretical analyses of question quality, there are some preliminary issues about questions that must be addressed: the contrast between presupposition and focus, the assumptions behind information-seeking questions, and an analysis of what constitutes a question. These issues and qualitative dimensions are briefly summarized in this section. More extensive discussions are available in other reports (Graesser, 1993; Graesser et al., in press; Graesser, Person et al., 1992).

#### Preliminary Theoretical Analyses of Questions

*Presupposition and focus.* A question can be decomposed into presupposed information and the focal information being queried. For example, in the question "What is the appropriate statistic to perform in the experiment?" there are at least two *presuppositions* (i.e., there was an experiment, someone performs an appropriate statistic) whereas the *focus* of the inquiry is on the type of statistic. The presupposed information in a discourse segment is in the common ground shared by the speaker and listener (Clark & Schaeffer, 1989). The focus of the question draws the answerer's attention to information that the questioner needs and hopes the answerer will supply.

*Assumptions behind information-seeking questions.* Speech act categories are normally defined according to the assumptions shared by speech participants rather than by syntactic or semantic regularities alone (Allen, 1983; Gibbs & Mueller, 1988; Searle, 1969). When a genuine information-seeking question is asked, the questioner is missing information and believes the answerer can supply it. Van der Meij (1987) identified several assumptions that must be met before

an utterance constitutes a genuine information-seeking question: (a) the questioner does not know the information asked for in the question, (b) the questioner believes that the presuppositions of the question are true, (c) the questioner believes that an answer exists, (d) the questioner wants to know the answer, (e) the questioner can assess whether a reply constitutes an answer, (f) the questioner believes that the answerer knows the answer, (g) the questioner believes that the answerer will not give an answer in absence of the question, (h) the questioner believes the answerer will supply the answer, (i) the questioner believes that the benefits of asking the question exceed the costs. A question fails to be an information-seeking question to the extent that these assumptions are violated. For example, some interrogative utterances are indirect requests for the listener to do something on behalf of the speaker, as in "Could you finish the session in the next five minutes?" (Clark, 1979; Gibbs & Mueller, 1988; Gordon & Lakoff, 1972). Teacher questions are usually not information-seeking questions because they violate assumptions a, d, f, and h. Most student questions are information-seeking questions, although it is important to acknowledge that student questions may reflect auxiliary motives, that is, to impress the teacher by asking a sophisticated question, to pass the burden of conversation onto the tutor. A speech act is a *bona fide inquiry* to the extent that all of the nine previously mentioned assumptions are met.

*What is a question?* There is the pressing issue of what constitutes a question, given that questions cannot be defined according to syntactic or semantic criteria alone. For the purposes of this article, a question is defined as a speech act that is either an inquiry (as defined previously), an interrogative expression, that is, an utterance that would be followed by a question mark in print, or both. The following expressions are inquiries, but only the first is an interrogative expression: "What is a factorial design?" (interrogative mood), "Tell me what a factorial design is" (imperative mood), and "I need to know what a factorial design is" (declarative mood). All of the following expressions are in the interrogative mood, but only the first is an inquiry: "What is a factorial design?" (inquiry), "Could you stop the session in 5 minutes?" (indirect request), and "Why did I ever take this course?" (gripe). Therefore, there is not a simple mapping between the syntactic mood of an utterance and its pragmatic speech act category. In an effort to be inclusive in our analysis of questions, all of these types of expressions would be counted as questions.

### Theoretical Dimensions That Address Question Quality

*Content of information sought.* Table 1 presents the 18 question-content categories in the GPH scheme (Graesser, Person et al., 1992). These categories are defined according to the content of the information sought rather than on the question stems (i.e., why, how, where, etc.) For example, *antecedent* questions tap previous events and states that cause or enable an event to occur. Antecedent questions can be articulated linguistically with a variety of stems: "Why did the event occur?" "How did the event occur?" "What caused the event to occur?" and so on. Most of the questions have an interrogative syntactic form.

Table 1  
**Question Categories in Graesser, Person, and Huber (1992) Scheme**

Question category	Abstract specification	Example
Short answer		
Verification	Is a fact true? Did an event occur?	Is the answer 5?
Disjunctive	Is <i>X</i> or <i>Y</i> the case? Is <i>X</i> , <i>Y</i> , or <i>Z</i> the case?	Is gender or female the variable?
Concept completion	Who? What? What is the referent of a noun argument slot? What qualitative attributes does entity <i>X</i> have?	Who ran this experiment?
Feature specification	What is the value of a quantitative variable?	What are the properties of a bar graph?
Quantification	How many?	How many degrees of freedom are on this variable?
Long answer		
Definition	What does <i>X</i> mean?	What is a <i>t</i> test?
Example	What is an example label or instance of the category?	What is an example of a factorial design?
Comparison	How is <i>X</i> similar to <i>Y</i> ? How is <i>X</i> different from <i>Y</i> ?	What is the difference between a <i>t</i> test and an <i>F</i> test?

Interpretation	What concept or claim can be inferred from a static or active pattern of data?	What is happening in this graph?
Causal antecedent	What state or event causally led to an event or state?	How did this experiment fail?
Causal consequence	What are the consequences of an event or state?	What happens when this level decreases?
Goal orientation	What are the motives or goals behind an agent's action?	Why did you put decision latency on the y-axis?
Instrumental/procedural	What instrument or plan allows an agent to accomplish a goal?	How do you present the stimulus on each trial?
Enablement	What object or resource allows an agent to perform an action?	What device allows you to measure stress?
Expectational	Why did some expected event not occur?	Why isn't there an interaction?
Judgmental	What value does the answerer place on an idea or advice?	What do you think of this operational definition?
Assertion	The speaker makes a statement indicating he lacks knowledge or does not understand an idea.	I don't understand main effects.
Request/Directive	The speaker wants the listener to perform an action.	Would you add those numbers together?



The question-content categories vary in the length of the expected answers. Questions that invite short answers place few demands on the answerer because a satisfactory answer is only a word or phrase. For example, acceptable answers to a verification question are "Yes," "No," or "Maybe." Answers to long-answer questions typically involve several sentences. One way to stimulate a listener to talk is to ask long-answer questions. One might expect a good tutor to ask many long-answer questions to diagnose the students' patterns of reasoning, misconceptions, and amount of knowledge. An evasive student might ask the tutor long-answer questions to shift the burden of conversation onto the tutor.

Reasoning is at the heart of the answers in some question-content categories. In logical reasoning, the statements expressed in an answer consist of premises and conclusions in a logical syllogism. In causal reasoning, the answer conveys events and states in causal chains. In goal-oriented reasoning, the answer traces the goals and planning structure of agents. Some of the analyses in this study focus on *deep-reasoning questions*, which elicit patterns of reasoning in logical, causal, or goal-oriented systems. Such questions include the following categories in Table 1: antecedent, consequence, goal-orientation, instrumental/procedural, enablement, and expectational. These questions are manifested in a tutoring session to the extent that the tutor and student explore deeper levels of comprehension. These deep-reasoning questions are highly correlated with the deeper levels of cognition in Bloom's taxonomy of educational objectives in the cognitive domain (Bloom, 1956). Good students and tutors are expected to ask these deep-reasoning questions.

Some questions are hybrids of two or more question categories. For example, verification questions are frequently combined with another category. The question "Did the drug dosage decrease the anxiety?" is an amalgamation of a verification question and an antecedent question. The fact that there are hybrid questions should not be construed as a weakness in the classification scheme. Most of the adequate classification schemes in the social sciences are *polythetic* rather than *monothetic* (Stokal, 1974). Each observation can be assigned to one and only one category in a monothetic classification scheme, whereas an observation can be assigned to multiple categories in a polythetic classification.

*Question-generation mechanisms.* The GPH scheme specifies four major mechanisms that generate questions in naturalistic conversation (including tutorial dialogue). Each major mechanism is segregated into subtypes, but these subtypes will not be contrasted in this article. An analysis of question-generation mechanisms incorporates the discourse context of a question. Judges must carefully analyze the goals, plans, and knowledge of the speech participants when identifying the mechanisms that elicit or motivate a question. Moreover, these question-generation mechanisms should be construed as being orthogonal to the category scheme in Table 1. For example, a verification question could in principle be motivated by any of the four major question-generation mechanisms.

The first major question-generation mechanism consists of information-seeking questions that occur when the questioner detects a *knowledge deficit* in his or her own knowledge base. These occur under the following condi-

tions: (a) when the questioner encounters an obstacle in a plan or problem, (b) when the questioner encounters a contradiction, (c) when an unusual or anomalous event is observed, (d) when there is an obvious gap in the questioner's knowledge base, and (e) when the questioner needs to make a decision among a set of alternatives that are equally likely. Good students presumably ask these knowledge-deficit questions to the extent that they are capable of self-regulating their knowledge.

The example knowledge-deficit question that follows (in italics) was extracted from a tutoring session on factorial designs in a college research methods course.

*Tutor:* Cells are also the same thing as groups or called experimental conditions. So those little boxes could be called cells, they could be called groups, they could be called experimental conditions.

*Student:* Wait a minute. *When you say "boxes" what do you mean?*

This student has a gap in her knowledge base about "boxes," so she asks the tutor to clarify the meaning of that term. Regarding the question-content categories in Table 1, this would be a definitional question; regarding the classification of questions on question-generation mechanisms, the question would be considered a knowledge-deficit question (and also a common-ground question). The following example was also extracted from a tutoring session on factorial designs.

*Tutor:* Is there a main effect for "A"?

*Student:* I don't think so.

*Tutor:* You don't think so?

*Student:* (Laughs.) *Is there one?*

This student spotted a contradiction between his own belief that there was no main effect of A and the tutor's expressed doubt or uncertainty about the main effect. The student asked the question to resolve the apparent contradiction. The question would be classified in the knowledge-deficit category with respect to question-generation mechanism; regarding its question-content category, the question would be a hybrid between a verification and an interpretational question.

The second major question-generation mechanism monitors the *common ground* between speech participants. The speech participants need to establish, negotiate, and update their mutual knowledge to achieve successful communication (Clark & Schaefer, 1989). Questions are generated to inquire whether the listener knows anything about a topic (e.g., "Have you covered factorial designs?"), to verify that a belief is correct (e.g., "Doesn't a factorial design have two independent variables?"), and to gauge how well a listener is understanding (e.g., "Do you understand?"). A good tutor should be able to identify the common ground between tutor and student and to expand the student's knowledge at the fringe or frontier (Brown & Burton, 1978).

The following excerpt was extracted from a tutoring session on variables in a research methods course.

*Tutor:* Suppose I am studying the effects of dehydration on the exploratory behavior of rats. One rat has had 23 hours of deprivation while another has had 30. What is the variable in that situation? What are some of the variables?

*Student:* Did you say dehydration?

The student asked this “counterclarification” question to confirm that the tutor had used a particular term (*dehydration*). Explicit utterances are supposed to be automatically included in the common ground of speech participants (Clark & Schaefer, 1989), so it is important that a listener registers what is said. Sometimes common-ground questions address inferences or intended meanings rather than explicit information, as illustrated in the following.

The following excerpt was extracted from a tutoring session on interactions in factorial designs.

*Tutor:* Go ahead and put that one up on the board. I’ll go ahead and give you the numbers.

*Student:* Just the graph?

The subtypes of common-ground questions are quite different for students and tutors. Students frequently want to confirm that their own beliefs are correct, so they ask common-ground questions that get the tutor to confirm or disconfirm such beliefs (e.g., Doesn’t the dependent variable go on the vertical axis?). In contrast, tutors frequently ask common-ground questions that assess what the student knows. For example, when the tutor asks the student “Do you know what an antagonistic interaction is?”, the tutor already knows the answer and is assessing the student’s amount of knowledge about the topic. This is a common-ground question rather than a knowledge-deficit question. Tutors frequently ask common-ground questions that inquire whether the student is understanding the material at a global level (e.g., “Do you understand?” “Are you following?”).

The third major question-generation mechanism coordinates social actions among speech participants. These questions are needed for multiple agents to collaborate in group activities and for single agents to get other agents to do things. These questions include indirect requests (e.g., “Would you do *X*?”), indirect advice (“Why don’t you do *X*?”), permission (“Can I do *X*?”), offers (“Can I do *X* for you?”), and negotiations (“If I do *X*, will you do *Y*?”). Tutors sometimes ask these *social-coordination* questions to engage students in activities that have pedagogical significance (e.g., “Could you graph these data on the board?”, “Why don’t you compute that test?”).

The fourth question-generation mechanism includes questions that are asked to control the flow of conversation and the attention of speech participants. These *conversation-control* questions include greetings, gripes, replies

to summons, rhetorical questions, and questions that change the flow of conversation. These questions were quite rare in the tutoring sessions.

*Degree of specification.* Questions vary on the degree to which the linguistic content specifies the information being sought. Questions with high specification have words or phrases that refer to the elements of the desired information and the relevant presupposed information. Questions with low specification have few words and phrases, so a dialogue context is needed for the answerer to fill in the missing information. A question is misinterpreted when the question has low specification and the answerer does not understand the dialogue context. An example high-specification question is "What are the variables in the factorial design in Experiment 1?"; the same question could be expressed as a low-specification question, for example, "What are those?" A good tutor should pose questions that have a sufficiently high specification for the student to understand them.

### **Tutoring Sessions Investigated in This Study**

The questions in two samples of tutoring sessions were analyzed in this study. Both the frequency and the qualitative characteristics of the questions were examined, using the theoretical scheme of Graesser, Person et al. (1992). Although analyses were conducted on both the tutor questions and the student questions, we were particularly interested in the student questions because they reflect active learning and they address the concern articulated earlier that student questions in a classroom are both infrequent and unsophisticated.

The primary sample of tutoring sessions consisted of upper-division college students in a research methods course. The tutoring sessions covered difficult topics in the research methods course, particularly those that involve quantitative reasoning. The tutoring sessions were a course requirement and counted as 6% of the final grade. The tutors were graduate students who had received As in a graduate-level research methods course. These tutors were not research assistants in the course and were not aware of the research that was being conducted on the tutoring sessions.

There are several considerations that motivated our selection of this tutoring sample. First, we desired a tutoring situation in which tutoring is known to be comparatively effective. According to available studies of tutoring (Cohen et al., 1982; Fitz-Gibbon, 1977), tutoring is more effective on topics that involve quantitative skills than on nonquantitative topics, and tutoring is more effective when it is highly integrated with the course curriculum than when it is an extra activity. Second, we desired a sample of tutors that were reasonably representative of the tutors in normal tutoring environments. Tutors are normally older students, paraprofessionals, and adult volunteers who have *not* been extensively trained in tutoring techniques (Fitz-Gibbon, 1977). Third, we desired a representative sample of college students rather than a restricted sample of students who were having difficulties. Given that all students in the course participated in the tutoring sessions, there was a wide range of abilities and levels of achievement within the college population.

Examination scores were available for the undergraduate students, so we could investigate the relationship between academic achievement in the course and indices of student question asking. We could assess, for example, whether more questions are asked by good students (who presumably self-regulate their learning) or bad students (who have more knowledge deficits). We could assess whether good students ask a higher proportion of high-quality questions, such as deep-reasoning questions. As mentioned earlier, however, it was beyond the scope of this study to investigate whether indices of student questioning and tutor questioning have a causal impact on learning outcomes.

A second sample of tutoring sessions was collected to provide some assessment of the generality of the results from the first sample. The second sample included seventh graders who were having problems with algebra. Specifically, the sample included almost all of the tutoring sessions in algebra that were arranged at a local school during a one-month time span. Examination scores were not available for these students, but we were able to examine the characteristics of the questions and to compare these characteristics with those of college students. We selected this second sample because it also involved tutoring on quantitative reasoning skills and because the students were old enough to be verbally articulate.

## Method

### Sample 1: College Students Learning About Research Methods

*Students and tutors.* Tutoring protocols were collected from 27 undergraduate students at Memphis State University. The students were enrolled in an upper division laboratory psychology course on research methods in the behavioral sciences. All students completed the tutoring sessions to fulfill a course requirement (6% of the total points in the course). Therefore, we had tutoring protocols on a representative sample of college students taking the class, as opposed to a restricted sample of students who were having difficulty with the material.

Three psychology graduate students were selected as tutors. Each tutor had received an A in an undergraduate research methodology course and a graduate-level research methodology course. Each tutor had tutored students on a few occasions prior to this study, but not in the area of research methods. Therefore, the tutors had a modest amount of tutoring experience, but they did not have extensive tutoring experience and did not receive training on the tutoring process. As discussed earlier, these characteristics are representative of the tutors that are available in most school systems; that is, typical tutors have moderately high domain knowledge, a small or moderate amount of tutoring experience, and minimal training on the tutoring process. Each tutor was paid \$500 for serving as a tutor in 18 tutoring sessions. The tutors were not teaching assistants for the course and were not research assistants in this study. They also were not informed about the purpose of the research project and the analyses that would be conducted.

*Learning materials.* The course instructor selected six topics that were normally troublesome for students in the course. Each topic had related sub-topics that would be covered in the tutoring session. An index card was prepared for each of the six topics; subtopics were listed with each topic, as specified in the following.

- *Variables:* operational definitions, types of scales, values of variables
- *Graphs:* frequency distributions, plotting means, histograms
- *Statistics:* decision matrix, Type I and II errors, *t* tests, probabilities
- *Hypothesis Testing:* formulating a hypothesis, practical constraints, control groups, design, statistical analyses
- *Factorial Designs:* independent variables, dependent variables, statistics, main effects, cells, interactions
- *Interactions:* independent variables, main effects, types of interactions, statistical significance

Students were exposed to the material covered on a topic before they participated in a tutoring session. This was accomplished in two ways. First, each topic was covered in a classroom lecture by the instructor before that topic was covered in a tutoring session. Second, both the tutors and students were required to read specific pages in the research methods text for the course (entitled *Methods in Behavioral Research*, Cozby, 1989) prior to the tutoring session. A mean of 14 pages was read prior to a tutoring session. Given that the students already had substantial exposure to the tutoring topics, it is safe to assume that they had an elementary to moderate amount of knowledge about each topic.

The tutoring sessions spanned an 8-week period. The topics covered during the first 3 weeks were variables, graphs, and statistics, with one topic covered per week. A 2-week break followed the first three tutoring sessions. The remaining three topics were covered during the subsequent 3 weeks.

*Equipment and setting.* The room used for the tutoring session was equipped with a videocamera, a television set, a marker board, colored markers, and the Cozby textbook. The television screen was covered during the entire sessions. The camera was positioned so that the student and the entire marker board was in sight. Therefore, the transcripts of the tutoring sessions included both spoken utterances and messages on the marker board.

*Procedure.* Prior to the tutoring session, the students were told that they would receive tutoring on particular pages in the Cozby text. When a student entered the tutoring room, the student was instructed to sit in view of the camera and to read the topic card aloud. The topic card outlined the material in the designated readings. The tutoring session subsequently proceeded in a natural fashion, in whatever direction the tutor and student saw fit. The three tutors were not given a specific format to follow, but they were told to resist the temptation of simply lecturing. The tutoring session lasted approximately 60 minutes, ranging between 45 minutes (a minimum) and 80 minutes.

Each of the 27 students participated in two tutoring sessions. This permitted us to assess whether student questioning changed when the students had

a small amount of exposure to the tutoring process (i.e., by comparing Session 1 with Session 2). A counterbalancing scheme was designed so that (a) a student never had the same tutor twice, (b) each tutor covered all six topics, (c) each tutor was assigned to 18 tutoring sessions, and (d) a student was tutored once during the first 3 weeks (i.e., three topics) and once during the last 3 weeks. Therefore, each tutor had three students assigned to each of the six tutoring topics, which yielded 54 sessions altogether. Ten of the 54 tutoring sessions could not be transcribed because of audio problems or extreme video problems that made it difficult to decipher messages on the marker board.

*Examination scores.* Achievement was measured by a total examination score, based on three objective examinations during the course. There was a total of 150 four-alternative, forced-choice questions in the total examination score. Most of the items (67%) were selected by the instructor from an item pool in an Instructor's Manual associated with the Cozby text. The other items were prepared by the instructor of the course. All 150 items had been prepared by the instructor before the semester in which the tutoring sessions were collected. The mean composite examination score of the 27 students was 100.6 ( $SD = 11.4$ ). Given that the students answered 67% of the questions correctly ( $100.6 \div 150 = .67$ ), the examination did not suffer from either a ceiling effect or floor effect. The tutors were not informed about the content and composition of the examinations.

### Sample 2: Seventh Graders Learning About Algebra

*Subjects and tutors.* We collected 22 tutoring sessions at a local middle school in Memphis. There were 13 seventh-grade students who were having trouble with particular topics in their algebra class (according to the teachers). The tutors were 10 high school students who normally provided these tutoring services for the middle school. On the average, a tutor had 9 hours of prior tutoring experience before tutoring a student in this sample. The sample of tutoring sessions included almost all of the tutoring sessions on algebra that occurred in the middle school for seventh graders during a one-month period. All of the tutoring sessions were naturally occurring tutoring sessions at the middle school. Unlike the college sample, the tutoring sessions in this sample were remedial activities and were not course requirements.

*Tutoring topics and sessions.* Almost all of the tutoring sessions covered three tutoring topics frequently problematic to seventh graders. These include: (a) exponents, (b) constructing equations from algebra word problems, and (c) fractions. A chapter or chapter section from a textbook was normally associated with each topic. Tutors and students frequently referred to this material during the tutoring sessions. The tutoring sessions lasted approximately 60 minutes, which was comparable to that of the research methods tutoring sessions. A research assistant from Memphis State University videotaped the sessions in a manner similar to the videotapings for the college sample.

### Transcription and Coding of the Tutoring Sessions

Transcribers received a one-hour training session on how to transcribe the pro-

ocols. They were instructed to transcribe the entire tutoring sessions verbatim, including all “ums,” “ahs,” word fragments, broken sentences, and pauses. The transcribers specified whether the tutor or the student spoke during each turn of the exchange. The transcribers sketched any messages that had occurred on the marker board in as much detail as possible. Hand gestures were specified in parentheses whenever the student or tutor pointed to the marker board or the book. Whenever the tutor and student spoke simultaneously, the parallel speech streams were placed within brackets. Each written transcription was verified for accuracy by a research assistant who spot-sampled random segments in the session.

Six trained judges coded the questions in the transcripts on a number of dimensions that were described in the Introduction. These judgments were made in the context of the tutorial dialogue rather than speech acts in isolation. The judges were graduate and undergraduate research assistants.

*Question identification.* Two judges achieved a high reliability score in deciding whether or not a speech act was a question, as defined in the Introduction (Cronbach's alpha = .96 or higher).

*Degree of specification.* Two judges categorized each question as to whether it had a high, medium, or low degree of specification. They achieved a spot-sample reliability score of .94.

*Question-content category.* Two judges were trained to classify the questions on the question-content categories in Table 1. For each question, the judges independently decided which of the 18 categories best fit the question. If a judge believed that a second or third category was also appropriate, the judge specified the second or third category assignments. The judges achieved a spot-sample reliability score of .96 or higher in deciding whether a particular category would be assigned to a question. Approximately half of the questions were assigned to only one question category (51%), whereas less than 1% of the questions were hybrids of three content categories.

The question-category analysis permitted us to use either a monothetic or a polythetic classification scheme. In the monothetic scheme, the categories were mutually exclusive, so we had to formulate a set of priorities whenever a question was a hybrid of two or more categories. There were two specific rules for assigning priority. First, the verification, assertion, and request/directive categories had lower priority than the other 15 question categories. The rationale for this priority rule is that the latter 15 categories were more specific regarding the content of the information sought. Second, short-answer questions had lower priority than long-answer questions. The rationale for this rule is that reasoning patterns are uncovered primarily in long-answer questions and there was theoretical interest in uncovering questions that tap deep reasoning. The most frequent hybrid by far was the amalgamation of a verification question with some other category. For example, the question “Is the mean of the sample 4.5?” is a hybrid between a verification question and a quantification question. Verification questions receive lower priority than the other content categories, so this question would be assigned to the quantification category in the monothetic analyses. In the polythetic scheme, each question could be



assigned to one, two, or three question categories. The analyses reported in the Results section focused exclusively on the monothetic classification scheme.

*Question-generation mechanism.* Two judges were trained to assign questions to the four question-generation mechanisms described in the Introduction: knowledge deficit, common ground, social coordination, and conversational control. Each question could be assigned to one or two categories. Most of the questions (72%) were assigned to only one category. The two judges achieved a satisfactory reliability score (Cronbach's alpha = .81 or higher) when deciding whether a mechanism was assigned to a question. In the case of hybrid questions in monothetic analyses, the knowledge-deficit category had higher priority than the other three categories because we were interested theoretically in the extent to which the subjects took an active role in self-regulating their knowledge. Also, the common-ground category had higher priority than the social-coordination and conversational-control categories (which were comparatively low in frequencies).

## Results

This section begins by reporting the overall number of questions asked by students and tutors in the tutorial sessions. The frequencies of these questions during tutoring were contrasted with those in classroom studies. The qualitative characteristics of questions were subsequently examined by analyzing the questions on three dimensions: question category, question-generation mechanism, and degree of specification. In all analyses, results from the two samples of tutoring sessions were compared (i.e., the Research Methods sample versus the Algebra sample). Finally, achievement scores of students in the Research Methods sample were correlated with frequencies of student questions and qualitative characteristics of student questions.

The student-tutor dyad (i.e., single tutoring session) was treated as the unit of analysis in all tests of statistical significance. Therefore, Tutor *A* with Student *X* was regarded as a separate case than Tutor *B* with Student *X*; similarly, Tutor *A* with Student *X* was regarded as a separate case than Tutor *A* with Student *Y*. By treating the dyad (or tutoring session) as the unit of analysis, we assume that the unique conversational patterns between a particular tutor and student are more fundamental than consistent patterns of a tutor across different students, and consistent patterns of a student across tutors. There is some precedence in treating the dyad as the functional unit of analysis in conversation research (Kenny, 1990) because speech participants mutually constrain each other in conversation. Statistical tests involved a between-group variable rather than a repeated-measures variable in all comparisons between tutors and students.

### Overall Number of Student and Tutor Questions

The mean number of student questions per tutoring session in the Research Methods sample was 21.1 ( $SD = 13.0$ ), whereas the mean in the Algebra sample was 32.2 ( $SD = 19.7$ ). These means were compared to the number of stu-

dent questions in classroom settings, which have been extensively documented in the literature. As discussed in the Introduction, the collective number of student questions in a classroom (per hour, approximately) ranged from 1.3 to 4.0, with a median of 3.0. The mean number of student questions in the tutoring sessions (which also lasted approximately one hour) exceeded the upper bound of 4.0 student questions in the classroom  $t(43) = 8.74, p < .05, S_e = 1.96$  in the Research Methods sample;  $t(21) = 6.91, p < .05, S_e = 4.09$  in the Algebra sample. These data support the general conclusion that student questions are more prevalent in a tutoring environment than in a classroom.

The 3.0 estimate of student questions in a classroom consists of the total number of questions asked by all of the students in a classroom. Given that an average classroom has approximately 26.7 students (Dillon, 1988), the mean number of questions asked by a particular student is only .11 question per hour. We verified this rate by tape-recording 12 hours of classroom lectures in the research methods course, focusing on those hours that covered the same topics as the tutoring sessions in the Research Methods sample. The rate of student questions was .17 question per student in the classroom, which compares favorably with the .11 estimate from the literature. Given that a particular student asked an average of 26.5 questions per hour in the tutoring sessions in the two samples of this study, and given that the classroom questioning literature estimates that a particular student asks .11 question per hour in a classroom setting, the incidence of student questions during tutoring is approximately 241 times the incidence of student questions in classroom settings (from the perspective of a single student).

The mean numbers of tutor questions per hour were 95.2 in the Research Methods sample ( $SD = 60.9$ ) and 112.1 in the Algebra sample ( $SD = 51.7$ ) These rates are significantly higher than the 69 question estimate of teacher questions in classrooms,  $t(43) = 2.85, p < .05, S_e = 9.2$  for the Research Methods sample and  $t(21) = 3.91, p < .05, S_e = 11.0$  for the Algebra sample. It follows that the incidence of tutor questions in the tutoring environment is 1.5 times the incidence of teacher questions in a classroom.

According to the previously mentioned data, tutor questions are much more prevalent than student questions in the tutoring sessions, just as teacher questions are more prevalent than student questions in the classroom. The differences between tutor and student questions were significant in both the Research Methods sample,  $F(1, 86) = 62.28, p < .05, MS_e = 1936.9$ , and in the Algebra sample,  $F(1, 42) = 46.13, p < .05, MS_e = 1520.8$ . The frequency scores indicate that 80% of questions in a tutoring session are asked by the tutor (82% in the Research Methods sample and 78% in the Algebra sample). This percentage is somewhat lower than the percentage of teacher questions in a classroom, which is estimated at 96% (i.e.,  $69 \div 72$ ).

The finding that student questions are more prevalent in one-to-one tutoring than in classroom settings should be moderately encouraging to educators who seek educational environments that promote active learning. Students apparently have more opportunity to self-regulate their learning in tutoring environments by asking more questions. According to the active injury hypothesis,

the increase in student question asking may explain why learning is better in tutoring environments than in classrooms. According to alternative hypotheses, however, it is important to consider the quality of questions and the quality of pedagogical activities. Subsequent analyses examined the quality of the questions generated by students and tutors.

### Question-Content Categories

The questions in the two samples were segregated into the 18 question-content categories (see in Table 1). There were 5,117 questions in the Research Methods sample and 3,174 questions in the Algebra sample, so the question corpus was quite large. Table 2 presents the breakdown of tutor questions and student questions in each sample. We adopted a monothetic classification scheme so the proportion scores in each column add to 1.00 (except for round-off error). A polythetic classification is not reported because it did not add to or alter any of the conclusions that will be made on the basis of the monothetic classification.

The distributions of proportion scores among the 18 question categories were quite similar for the two samples of tutoring sessions. We performed correlations on the proportion scores presented in Table 2. There was a high and significant correlation between the set of proportions of tutor questions in the Research Sample and that of the Algebra sample,  $r = .95$ ,  $p < .05$ ; the comparable correlation was also high and significant for student questions,  $r = .61$ ,  $p < .05$ .

Verification questions were the most prevalent questions for tutors (.45). There were moderately high proportion scores for concept completion questions (.10), quantification questions (.09), and instrumental-procedural questions (.09) in the case of tutors. It should be noted that short answers are appropriate for most of these question categories. In the case of students, the order of the top four categories was verification (.22), instrumental-procedural (.21), concept completion (.11), and quantification (.11). Therefore, the same four categories emerged as being most frequent for both tutors and students.

The question distributions indicated that tutors asked a higher proportion of short-answer questions than did students. As designated in Tables 1 and 2, short-answer questions are segregated from long-answer questions; the assertion and request/directive categories are not included because they are ambiguous with respect to answer length. Analyses of Variance (ANOVAs) were performed on proportion scores in a factorial design that contrasted length of expected answer (short-versus long-answer question categories) and speech participant (student versus tutor). In the Research Methods sample, short-answer questions were more prevalent than long-answer questions, .56 versus .38, respectively,  $F(1, 84) = 12.94$ ,  $p < .05$ ,  $MS_e = .06$ , but there was no significant length  $\times$  speech participant interaction. The difference between short- and long-answer questions appears to be more pronounced for tutors (.60 versus .35) than for students (.52 versus .40), but there was no significant interaction. In the Algebra sample, once again the short-answer questions were more prevalent than long-answer questions, .60 versus .38, respectively,  $F(1, 42) = 3.47$ ,  $p < .05$ ,  $MS_e = .03$ , but there was a length  $\times$  speech participant interac-

Table 2  
**Question Categories of Questions That Students and Tutors Ask**

Question category	Research methods (college students)		Algebra (7th graders)	
	Student	Tutor	Student	Tutor
<b>Short-answer questions</b>				
1. Verification (Yes/No answer)	.28	.43	.15	.47
2. Disjunctive (A or B?)	.02	.03	.02	.01
3. Concept completion (Who? What?)	.14	.09	.08	.10
4. Feature specification	.05	.02	.03	.01
5. Quantification (How many?)	.03	.03	.18	.14
<b>Long-answer questions</b>				
6. Definition (What does X mean?)	.03	.05	.00	.01
7. Comparison (How does X compare to Y?)	.02	.02	.01	.00
8. Example (What is an example of X?)	.01	.01	.00	.01
9. Interpretational (What's happening?)	.09	.09	.08	.05
10. Judgmental	.03	.02	.01	.02
11. Antecedent (Why?)	.02	.02	.01	.01
12. Consequence (What if?)	.03	.04	.05	.04
13. Goal orientation (Why?)	.01	.02	.00	.00
14. Enablement (Why/How?)	.00	.01	.01	.00
15. Instrumental/procedural (How?)	.13	.06	.29	.11
16. Expectational (Why not?)	.03	.01	.03	.01
<b>Other</b>				
17. Assertion	.08	.00	.07	.00
18. Request/directive	.03	.05	.01	.04

tion,  $F(1, 42) = 14.41, p < .05, MS_e = .03$ . The difference between short- and long-answer questions was not significant for students (.46 versus .49) whereas it was robust for tutors (.73 versus .26). The tutors tended to place minimal demands on the student because the tutors asked comparatively few questions that require long-winded answers (compared to the questions asked by students). In contrast, students either encouraged or permitted the tutor to talk because they asked a comparatively high proportion of long-answer questions. The students perhaps attempted to place the burden on the tutor to articulate difficult material and to trace the steps of reasoning.

Proportion scores were examined for the deep-reasoning questions, that is, why, why not, how, what if. These question categories include antecedent, consequence, goal-orientation, enablement, instrumental-procedural, and expectational questions. These proportion scores were significantly higher for students than for tutors in the Research Methods sample, .22 and .16, respectively,  $F(1, 84) = 5.58, p < .05, MS_e = .02$ , and in the Algebra sample, .39 versus .17,  $F(1, 42) = 20.69, p < .05, MS_e = .02$ . Once again, the tutors were comparatively reluctant to ask these deep-reasoning questions and thereby impose demands on students (compared to the questions asked by students). Viewed differently, the students were more likely to select questions that expose the tutor's reasoning and problem solving.

### Question-Generation Mechanisms

The questions were segregated into the four question-generation mechanisms that were defined in the Introduction: knowledge deficit, common ground, social coordination, and conversational control. Proportion scores are presented in Table 3 for the two samples of tutoring sessions.

The two samples of tutoring sessions showed similar patterns of proportion scores. We performed an ANOVA on proportion scores as a function of speech participants (tutor versus student) and two question-generation mechanisms (knowledge deficit versus common ground). The interaction terms were statistically significant in both samples,  $F(1, 86) = 134.14, p < .05, MS_e = .02$  for Research Methods and  $F(1, 42) = 45.34, p < .05, MS_e = .03$  for Algebra. The proportion scores for knowledge-deficit questions were higher

Table 3  
Question-Generation Mechanisms

Question-generation mechanism	Research methods (college students)		Algebra (7th graders)	
	Student	Tutor	Student	Tutor
Knowledge deficit	.33	.02	.24	.00
Common ground	.60	.90	.72	.93
Social coordination	.04	.02	.03	.04
Conversational control	.03	.06	.01	.03

for students than for tutors in the Research Methods sample,  $F(1, 86) = 166.93$ ,  $p < .05$ ,  $MS_e = .01$ , and in the Algebra sample,  $F(1, 42) = 53.55$ ,  $p < .05$ ,  $MS_e = .01$ . In contrast, proportion scores for common-ground questions were lower for students than for tutors in the Research Methods sample,  $F(1, 86) = 86.91$ ,  $p < .05$ ,  $MS_e = .01$ , and in the Algebra sample,  $F(1, 42) = 35.24$ ,  $p < .05$ ,  $MS_e = .02$ . The other two question-generation mechanisms were comparatively rare, so statistical analyses are not reported.

It is informative that 29% of the student questions were attempts to correct knowledge deficits. It suggests that students somewhat take an active role in self-regulating their learning. Most of the knowledge-deficit questions ( $.23 \div .29 = 79\%$ ) were triggered by contradictions or by anomalous facts/events the students had trouble explaining. Another 17% of the knowledge-deficit questions were triggered by an obvious gap in the student's knowledge base; for example, the tutor mentions the term *antagonistic interaction* and then the student asks "So what is an antagonistic interaction?"

Most of the tutor questions and student questions monitored the common ground between the speech participants. Not surprisingly, most of the efforts of both parties were devoted to identifying, diagnosing, and verifying the knowledge of the student, as opposed to the tutor. Most of the students' common-ground questions (89%) were attempts to confirm the validity of their own beliefs (e.g., "Doesn't a factorial design have two independent variables?"); only 4% of the tutors' common-ground questions were attempts to confirm tutors' knowledge. The vast majority of the tutors' common-ground questions (82%) were assessments of the students' knowledge (e.g., "Do you know what a factorial design is?", "Do you understand?"); only 3% of the students' common-ground questions were assessments of the tutors' knowledge. As expected, therefore, there was an asymmetry in knowledge tracking because the primary focus was on the knowledge base of the student. The tutor presumably was viewed by both parties as having a good command of the material.

### Degree of Question Specification

The questions were classified on degree of specification, with values of high, medium, and low. Table 4 presents proportion scores, segregating tutor and student questions in the two samples. The data from the two samples were strikingly similar. A very small proportion (.02) of the questions had a high degree of specification. The proportion scores for high-specification questions were not significantly different for tutors and students in the two samples. In the case of low-specification questions, however, the proportion scores were higher for tutors than students in the Research Methods sample,  $F(1, 86) = 6.00$ ,  $p < .05$ ,  $MS_e = .03$ , and in the Algebra sample,  $F(1, 42) = 5.36$ ,  $p < .05$ ,  $MS_e = .03$ . There is a plausible explanation for the comparatively high proportion of low-specification questions for tutors. They frequently attempted to gauge whether the student was understanding the material by asking questions such as "Do you understand?" or "Are you following?"; 35% of the tutors' questions were general comprehension-gauging questions, all of which are low-specification questions.

Table 4  
**Degree of Question Specification**

Source	Degree of specification		
	High	Medium	Low
Student questions			
Research methods	.03	.67	.30
Algebra	.02	.60	.39
Tutor questions			
Research methods	.03	.50	.47
Algebra	.01	.47	.52

**Correlations Between Student Questions and Examination Scores**

If good students are inquisitive, then we would expect a positive correlation between examination scores and number of questions. On the other hand, good students have fewer knowledge deficits, which would yield a negative relationship between examination scores and number of questions. We performed some correlation analyses on the Research Methods sample because examination scores were available. Separate correlations were computed for the first half of the tutoring sessions and the second half. Each student had been assigned to one tutoring session during the first half (i.e., consisting of 3 weeks, covering three tutoring topics) and to one tutoring session during the second half.

Table 5 presents correlations between examination scores and measures of student question asking. The correlation between examination scores and overall number of student questions was significantly negative during the first half of the course, but not during the second half. Therefore, students with more knowledge deficits (i.e., lower examination scores) asked more questions in the first half of the course, but the trend did not persist in the second half. Perhaps the good students learned to ask more questions after they had some exposure to tutoring: This would explain the low, nonsignificant correlation in the second half of the course. The low correlation might be explained by a U-shaped function, such that both poor students and good students ask more questions than students with an intermediate level of achievement. However, a curvilinear relationship was not found to be statistically significant in any analysis in this study that examined the relationship between student questioning and examination scores (including all of the measures in Table 5).

We computed follow-up correlation analyses to assess whether good students tended to ask more questions that involved deep reasoning or knowledge deficits. Proportion scores were computed by dividing the frequency of deep-reasoning questions (or alternatively, knowledge-deficit questions) by the total number of student questions. We would expect a significant positive correlation between the students' examination scores and the proportion of deep-reasoning questions (or knowledge-deficit questions). As shown in Table

Table 5  
**Correlations Between Examination Scores  
 and Measures of Question Asking**

Question-asking measures	First half of course	Second half of course
Total number of student questions	-.56*	-.12
Proportion of student questions that are deep-reasoning questions	.36**	.47*
Proportion of student questions that are classified as deep on Bloom's taxonomy	-.14	.46*
Proportion of student questions that address knowledge deficits	-.01	.35**

\*Significant at  $p < .05$ , two-tailed. \*\*Significant at  $p < .05$ , one-tailed.

5, this prediction was confirmed in the second half of the course but not the first half. In the second half, there was a significant positive correlation between examination scores and (a) the proportion of questions that involved deep reasoning and (b) the proportion of questions that involved knowledge deficits; in the first half, these two correlations were not statistically significant. These data are consistent with the claim that the question-asking skills of good students are enhanced after some exposure to tutoring. Unfortunately, the tutoring topics were different in the first and second halves of the course, so we are uncertain whether the differences mentioned previously between the first and second halves can be attributed to exposure to tutoring or to variations among tutoring topics.

We have assumed that deep-reasoning questions including question-content categories 11–16 in Table 2 (e.g., why, why not, how, and what if questions). Such questions tap the steps and rationale in logical reasoning, in problem-solving procedures, in plans, and in causal sequences. We performed an addition analysis on the student questions to assess the validity of the claim that these question categories are truly deeper questions than the other question categories. Specifically, two research assistants classified each question on depth by adopting Bloom's taxonomy of educational objectives in the cognitive domain (Bloom, 1956). There were five values on this scale, as specified as follows.

1. This lowest level of depth includes knowledge of specific terminology and specific facts (level 1.00 in Bloom's taxonomy). It includes student questions that repeat tutor questions, questions that seek clarification of the tutor's speech acts, and various questions that the judges considered shallow.
2. These questions involved ways and means of dealing with specifics, including knowledge of conventions, trends, classifications and categories, criteria, and methodology (Level 1.20 in Bloom's taxonomy).



3. These questions address knowledge of universals and abstractions in a field, including principles, generalizations, theories, and structures (Level 1.30 in Bloom's taxonomy).
4. These questions involve comprehension of the material to the extent the student can perform translations, interpretations, and extrapolations (Level 2.00 in Bloom's taxonomy).
5. These questions involve all of the deeper levels in Bloom's taxonomy (Levels 3.00 through 6.00): Application, analysis, synthesis, and evaluation.

The two judges could reliably classify the student questions on Bloom's taxonomy (Cronbach's  $\alpha = .85$ ).

Analyses revealed that the vast majority of student questions were at the low (shallow) end of Bloom's taxonomy. The distribution of questions among the five ratings was .70, .19, .04, .06, and .01 for ratings 1, 2, 3, 4, and 5, respectively. These data clearly indicate that student questions are rarely deep.

We conducted an analysis that assessed whether deep-reasoning questions were deep questions according to Bloom's taxonomy. For each student, we computed the proportion of student questions that were regarded as comparatively deep in Bloom's taxonomy (i.e., ratings of 2, 3, 4, or 5). We found that these proportions were highly correlated with the proportion of student questions that were deep-reasoning questions,  $r = .64$ ,  $p < .05$ . In contrast, the proportion of questions that were deep in Bloom's taxonomy was not significantly correlated with the proportion of knowledge-deficit questions ( $r = .24$ ) and common-ground questions ( $r = -.11$ ). Therefore, there is some validity to our claim that the deep-reasoning questions are truly deep questions. Moreover, according to Table 5, there was a positive correlation between student achievement and deep questions in Bloom's taxonomy: This trend occurred in the second half of the course but not in the first half, just as we found for deep-reasoning questions and knowledge-deficit questions.

## Discussion

The first objective of this study was to document the number of questions asked in one-to-one tutoring sessions and to compare these estimates to the questions in classroom studies. We found that a student in a tutoring session asked approximately 240 times as many questions during a tutoring as a particular student would ask in a classroom study. Tutors asked about 1.5 times as many questions as did teachers in a classroom setting. Therefore, tutoring clearly provides a setting for more active inquiry, particularly on the part of the student. Our estimates of questions during tutoring were based on two radically different samples of students, namely a representative sample of college students in a research methods class and a low-achieving sample of seventh graders attempting to learn algebra. The data were strikingly similar in these two samples of students. The empirical estimates for classroom questioning were based on studies that spanned several topics and countries (Buseri, 1988; Dillon, 1988; Fenclova, 1978; Good et al., 1987; Nickel & Fenner, 1974; Susskind, 1969),

including classroom lectures in a research methods course that covered the same material that was covered in the tutoring sessions of the college students.

We believe that our results are representative of tutoring in normal instructional settings, at least in the context of quantitative reasoning. However, more research on tutoring is clearly needed before we can offer general claims about tutoring that apply to diverse knowledge domains, age groups, educational programs, and so on. It is also possible that patterns of questioning will substantially vary as a function of the amount of tutoring experience that tutors and students receive. Additional research is needed to identify the qualitative patterns of discourse interaction in tutorial dialogue (Graesser, in press).

The previously mentioned findings support the claim that tutoring provides a social, cognitive, and pedagogical context for students to take more control over their learning and to correct their idiosyncratic knowledge deficits. Social barriers do not severely dampen inquisitiveness to the same extent as they do in a classroom. The "active inquiry" hypothesis provides one potential explanation of the finding that learning is superior in tutoring settings than in classroom environments (Bloom, 1984; Cohen et al., 1982). That is, students may learn more in tutoring sessions because they have more opportunities to ask questions that pertain to their knowledge deficits.

Computer software has recently been designed to permit extensive question asking by the learner; for example, a "Point and Query" (P&Q) system radically facilitates the speed and quality of questioning (Graesser, Langston, & Lang, 1992; Graesser, Langston, & Baggett, in press). The student learns entirely by asking questions and reading answers. To ask a question, the student first points to a word or picture element on the computer screen and then to a question that is relevant to the element (from a menu of relevant questions). The menu of relevant questions is formulated on the basis of background knowledge structures and a theory of human question answering called QUEST (Graesser & Franklin, 1990; Graesser & Hemphill, 1991; Graesser, Lang, & Roberts, 1991). The P&Q software is embedded in a hypertext system, so answers are preformulated and quickly accessed; the student can ask a question with two points of a finger (or two clicks of a mouse). The P&Q software is similar to some other menu-based question asking systems that have recently been developed (Schank, Ferguson, Birnbaum, Barger, & Greising, 1991; Sebrechts & Swartz, 1991).

Given that it is so easy to ask questions on these question-menu computer systems, one might inquire about the incidence of learner questions on these systems. Graesser, Langston, and Baggett (in press) collected data from college students who learned about woodwind instruments with the P&Q software. The goals of the learner were manipulated so that students were expected to acquire deep causal knowledge in one condition and superficial knowledge in another condition. The incidence of questioning was quite similar under these two conditions. Students asked approximately 135 questions per hour. Therefore, a student asks .11 question per hour in a classroom, 26.5 questions per hour in a tutoring setting with a human tutor, and 135 questions per hour in a computerized learning environment in which the only way to learn is to ask

questions. Students are clearly *capable* of engaging in active inquiry (as manifested by question asking), but the classroom environment does not foster it.

The reported tutoring data are consistent with the claim that students are to some extent capable of self-regulating their learning by asking questions when they spot knowledge deficits. We found that 29% of the students' questions addressed some form of knowledge deficit about the domain knowledge they were attempting to learn. These knowledge-deficit questions were triggered when the student identified a contradiction, an anomalous fact or event, or a word that was unfamiliar. Graesser and McMahan (1993) have also reported that anomalous information *causes* an increase in the incidence of question asking in experiments that have precise control over the stimulus material. In that study, subjects were instructed to generate questions while they solved different versions of quantitative problems: complete original, deletion of critical information, addition of a contradictory statement, addition of an irrelevant statement. The transformed versions of the problems produced more questions than did the original complete problem, and a subset of the questions addressed the transformations. Therefore, there is evidence that students can to some extent self-regulate their learning under both naturalistic tutoring environments and more controlled laboratory environments. At the same time, it is important to acknowledge that the students are not perfect in identifying their knowledge deficits. Students frequently miss contradictions and inconsistencies in scientific text, mathematical word problems, and other types of material (Baker, 1979; Glenberg, Wilkinson, & Epstein, 1982; Markman, 1979; Otero & Campanario, 1990).

Some additional findings are compatible with the claim that students have at least a modest ability to self-regulate their learning. First, 89% of the students' questions in the common-ground category were attempts to confirm that their own knowledge was correct. In essence, they were asking "Is it true that a particular belief that I have is correct?" and were seeking verification from the expert tutor. This indeed is an active monitoring of their own knowledge. Second, in the first half of the course there was a negative correlation between the students' examination scores and the overall number of questions that they asked. Thus, the students with more knowledge deficits ended up asking more questions. Other researchers have reported negative correlations between mastery of the material and the incidence of learner-questions in more restricted experimental tasks (Fishbein, Eckart, Lauver, van Leeuwen, & Langmeyer, 1990; Flammer, 1981). Third, there was some weak evidence that student achievement was positively correlated with the proportion of student questions that are knowledge-deficit questions. This positive relationship is compatible with the hypothesis that good students actively monitor their own comprehension failures (Brown et al., 1983; Chi et al., 1989; Zimmerman, 1989). However, this relationship was significant in the second half of the course, but not the first half. The good students apparently need some exposure to tutoring before they become sensitive to their knowledge deficits and ask questions that address these deficits. Unfortunately, the tutoring topics were different in the two halves of the course, so differences between the two halves could be attributed to

fluctuations among tutoring topics rather than to exposure to tutoring. Clearly, additional research is needed to assess how patterns of student questioning change as a function of tutoring experience.

The negative correlation between student achievement and overall number of questions seems incompatible with the hypothesis that good students actively monitor their own comprehension failures. We believe that the apparent contradiction can be resolved by considering the *quality* of the students' ability to self-regulate their knowledge and the *quality* of their questions. Specifically, good comprehenders acquire the knowledge more thoroughly, at deeper, more sophisticated levels; consequently, their questions are deeper, more sophisticated, and more focused on knowledge deficits. In contrast, the comparatively poor students ask a large number of unsophisticated questions.

Our analyses of student questions revealed that a comparatively small percentage of their questions were deep, sophisticated questions that penetrated the inherent complexity of the material. Whereas 92% of the student questions were at the shallow end of Bloom's educational objectives in the cognitive domain (namely level 1), only 8% spanned Levels 2 through 6. These results are consistent with the conclusion that most students have not completely mastered effective, sophisticated, question-asking skills.

Given that most students have not mastered effective question-asking skills, then there should be benefits in learning after they are taught how to ask good questions. In fact, there is substantial empirical evidence that there are robust improvements in the comprehension, learning, and memory of technical material after students are trained how to ask good questions (Davey & McBride, 1986; Gavelek & Raphael, 1985; King, 1989, 1990; Palincsar & Brown, 1984; Singer & Donlan, 1982). We should point out that the P&Q educational software has the potential to train students how to ask good questions. Students would learn the good questions for any given domain of knowledge (e.g., science, mathematics, history, etc.) to the extent that the good questions are included on the question menu (and bad questions excluded). The QUEST model currently specifies which categories of questions are good questions on the P&Q software. We would not be surprised if the QUEST model proves to be inadequate in discriminating good from bad questions, but it does provide a first step that is grounded both theoretically and empirically.

An important aspect of the tutoring process consists of monitoring the common ground between the tutor and student. Indeed, common ground is a pervasive problem in all forms of communication (Clark & Schaefer, 1989). The identification, establishment, and modification of common ground is a collaborative process between speech participants (Carlsen, 1991; Fox, 1988; Graesser, 1993, in press; Resnick, Salmon, & Zeitz, 1991). As expected, most of the attention was centered on the student's knowledge in these tutoring sessions. The majority of student questions were attempts to confirm his or her own beliefs. The majority of tutor questions were attempts to identify what the student knows. Available research on the process of tutoring indicates that tutors substantially vary in the extent to which they diagnose and repair a particular student's knowledge deficits and conceptual bugs (McArthur et al., 1990; Put-

nam, 1987). Some tutors grill students with a script of questions and with exercises that expose imperfect knowledge in a typical student (rather than a particular student). Other tutors are more sensitive to the idiosyncratic problems of a particular student.

Graesser (in press) reported that tutors and students frequently enact a five-step tutorial frame that establishes and updates common ground.

*Step 1: Tutor question:* The tutor asks the student a question to diagnose the student's knowledge about a subtopic, for example, "What is a factorial design?"

*Step 2: Student answer:* The student answers the tutor's question and thereby displays his or her knowledge about the subtopic, for example, "It has several variables."

*Step 3: Answer evaluation:* The tutor evaluates the quality of the student's answer, for example, "Well, not exactly."

*Step 4: Answer elaboration:* The answer is elaborated primarily by the tutor (and to some extent by the student) to improve the quality of the answer to the question, e.g., "A factorial design has at least two independent variables and one dependent variable."

*Step 5: Assessment of student's understanding:* The tutor assesses whether the student understands the answer by asking a comprehension-gauging question, "Do you understand?"

The normative rules of polite conversation are sometimes incompatible with the accurate establishment of common ground between tutor and student. For example, 35% of the tutors' questions were open-ended questions that gauged whether the student was understanding a topic being discussed (e.g., "Do you understand?" "Are you following?" "Okay?"). Students frequently answer "Yes" when they fail to understand the material because they want to be polite, because they want to avoid appearing ignorant, or because they are unable to detect their lack of understanding. Tutors often accept this feedback, assume the student has mastered the topic, and move on to another topic. Common ground is not accurately established when this occurs. In fact, Chi et al. (1989) have reported that the comparatively poor students tended to answer "Yes" to the question "Do you understand?" in tutoring sessions on physics; the good students could identify their own knowledge deficits and answer "No." Tutors and classroom teachers should be warned that it is inappropriate to trust the feedback that students provide about their own knowledge.

The problem of monitoring common ground is complicated further by the data on degree of question specification. Only 2% of the questions of both tutors and students had a high degree of specification. Half of the tutor questions (50%) and 35% of the student questions had a low degree of question specification. As the degree of question specification decreases, the listener has a higher likelihood of being confused and there is an increase in problems in establishing common ground. The context of the tutoring dialogue is frequently not sufficient to construct the intended meaning of a question when questions have a low or medium degree of specification (Graesser, 1993). The prevalence of

misunderstandings in dialogue has been documented in several contexts, including doctor-patient interactions, lawyer-witness interactions, and student-teacher interactions (Coombs & Alty, 1980; Edwards & Mercer, 1989; Labov & Fanshel, 1977; Valdez, 1986). Tutors should be instructed to formulate their questions with a higher degree of specification.

The tutors tended to ask short-answer questions that placed minimal burdens on the students in supplying information. For tutors, 67% of the questions were short-answer questions, 31% were long-answer questions, and 3% were in question categories that had an ambiguous status with respect to length of answer (i.e., assertions and requests/directives). In contrast, 49% of the student questions were short-answer questions and 45% were long-answer questions. Similarly, the proportion scores for deep-reasoning questions were higher for students than for tutors (.31 versus .17). Therefore, the students selected a comparatively high proportion of questions that invited lengthy answers and reasoning from the tutors, whereas tutors did not place comparable burdens on the student. This tendency is incompatible with the pedagogical goal of encouraging students to expose their reasoning and to organize coherent messages. Perhaps tutors should be instructed to ask more long-answer questions and more deep-reasoning questions.

The most prevalent question-generation mechanisms were attempts to correct knowledge deficits and to monitor common ground. The other two types of question-generation mechanisms were not very frequent because of the social constraints of the tutorial context. Tutoring does not ordinarily involve physical activities that span long time periods, so very few questions were attempts to coordinate social activity. Social-coordination questions (i.e., indirect requests, indirect advice, offers, permission, and negotiations) are more prevalent in contexts where there are several agents working together over a longer time span. Conversational-control questions were also infrequent because there are only two speech participants in a tutorial dialogue. Conversational-control questions are more frequent when several individuals are interacting socially and it is important to distribute contributions among many individuals.

This study has documented the questions asked during tutorial sessions and has discussed the relevance of question asking to theories of learning, cognition, and conversation. We have pointed out a few ways that tutors might improve their question-asking skills to facilitate student learning. Few studies have examined the process of tutoring and the impact of these processes on learning outcomes. A productive direction for future research will be to examine tutoring processes and strategies in greater detail and to determine which of these processes explain why tutoring is an excellent method of learning.

#### **Note**

This research was funded by grants awarded to the first author by the Office of Naval Research (N00014-88-K-0110, N00014-90-J-1492, and N00014-92-J-1826). We would like to thank Levy Eymard, John Huber, Brenda Johnson, Mark Langston, Joe Magliano, and John White for serving as judges in the question analyses. We are indebted to John Cady for providing access to the tutoring sessions for the seventh graders learning algebra.

## References

- Allen, J. (1983). Recognizing intentions from natural language utterances. In M. Brady & R. C. Berwick (Eds.), *Computational models of discourse* (pp. 107-166). Cambridge, MA: MIT press.
- Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science, 13*, 467-505.
- Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior, 11*, 363-374.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain*. New York: McKay.
- Bloom, B. S. (1984). The 2 Sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 4-16.
- Bransford, J. D., Arbitman-Smith, R., Stein, B. S., & Vye, N. J. (1985). Analysis—improving thinking and learning skills: An analysis of three approaches. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 1* (pp. 133-206). Hillsdale, NJ: Erlbaum.
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed.) New York: Wiley.
- Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science, 2*, 155-192.
- Buseri, J. C. (1988). Questions in Nigerian science classes. *Questioning Exchange, 2*(3), 275-280.
- Carlsen, W. S. (1991). Questioning in classrooms: A sociolinguistic perspective. *Review of Educational Research, 61*, 157-178.
- Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.
- Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology, 11*, 430-477.
- Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.
- Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248.
- Collins, A. (1985). Teaching and reasoning skills. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2* (pp. 579-586). Hillsdale, NJ: Erlbaum.
- Collins, A. (1988). Different goals of inquiry teaching. *Questioning Exchange, 2*(2) 39-45.
- Collins, A., Brown, J. S., & Larkin, K. M. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 385-407). Hillsdale, NJ: Erlbaum.
- Coombs, M. H., & Alty, J. L. (1980). Face-to-face guidance of university computer users. II: Characterizing advisory interactions. *International Journal of Man-Machine Studies, 12*, 407-429.
- Cozby, P. C. (1989). *Methods in behavioral research*. Mountain View, CA: Mayfield Publishing Company.
- Dahlgren, K. (1988). *Naive semantics for natural language understanding*. Boston: Kluwer Academic Press.
- Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology, 78*, 256-262.
- Dillon, J. T. (1984). The classification of research questions. *Review of Educational Research, 54*, 327-361.
- Dillon, J. T. (1988). *Questioning and teaching: A manual of practice*. New York: Teachers

College Press.

- Edwards, D., & Mercer, N. M. (1989). Reconstructing context: The conventionalization of classroom knowledge. *Discourse Processes*, 8, 229-259.
- Fenclova, J. (1978). How does a teacher of physics ask questions? *Mathematics and Physics at School*, 2(9), 134-137.
- Fishbein, H. D., Eckart, T., Lauver, E., van Leeuwen, R., & Langmeyer, D. (1990). Learners' questions and comprehension in a tutoring setting. *Journal of Educational Psychology*, 82, 163-170.
- Fitz-Gibbon, C. T. (1977). An analysis of the literature of cross-age tutoring. (ERIC Document Reproduction Service No. ED 148 807)
- Flammer, A. (1981). Towards a theory of question asking. *Psychological Research*, 43, 407-420.
- Fox, B. (1988). *Cognitive and interactional aspects of correction in tutoring* (Tech. Rep. No. 88-2). University of Colorado, Boulder, Colorado.
- Gavelek, J. R., & Raphael, T. E. (1985). Metacognition, instruction, and the role of questioning activities. In D. L. Forrest-Pressley, G. E. MacKinnin, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (Vol. 2, pp. 103-136). Orlando, FL: Academic Press.
- Gibbs, R. W., & Mueller, R. A. G. (1988). Conversational sequences and references for indirect speech acts. *Discourse Processes*, 11, 101-116.
- Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory and Cognition*, 10, 597-602.
- Good, T. L., Slavings, R. L., Harel, K. H., & Emerson, H. (1987). Student passivity: A study of question asking in K-12 classrooms. *Sociology of Education*, 60, 181-199.
- Gordon, P., & Lakoff, G. (1972). Conversational postulates. *Papers from the seventh regional meeting. Chicago Linguistics Society*, 7, 63-84.
- Graesser, A. C. (1993). Questioning mechanisms during complex learning. Memphis State University, Memphis, TN (ERIC Document Reproduction Service No. ED 350 306)
- Graesser, A. C. (in press). Dialogue patterns during naturalistic tutoring. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Graesser, A. C., & Franklin, S. P. (1990). QUEST: A cognitive model of question answering. *Discourse Processes*, 13, 279-303.
- Graesser, A. C., & Hemphill, D. (1991). Question answering in the context of scientific mechanisms. *Journal of Memory and Language*, 30, 186-209.
- Graesser, A. C., Lang, K. L., & Roberts, R. M. (1991). Question answering in the context of stories. *Journal of Experimental Psychology: General*, 120, 254-277.
- Graesser, A. C., Langston, M. C., & Baggett, W. B. (in press). Exploring information about concepts by asking questions. In G. V. Nakamura, R. M. Taraban, & D. Medin (Eds.), *Acquisition, representation, and processing of categories and concepts*. Orlando, FL: Academic Press.
- Graesser, A. C., Langston, M. C., & Lang, K. L. (1992). Designing educational software around questioning. *Journal of Artificial Intelligence in Education*, 3, 235-243.
- Graesser, A. C., & McMahan, C. L. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology*, 85, 136-151.
- Graesser, A. C., McMahan, C. L., & Johnson, B. (in press). Question asking and answering. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Graesser, A. C., Person N. K., & Huber, J. D. (1992). Mechanisms that generate questions. In T. E. Lauer, E. Peacock, & A. C. Graesser (Eds.), *Questions and information systems* (pp. 167-187). Hillsdale, NJ: Erlbaum.
- Hilton, D. J. (1990). Conversational processes and causal explanation. *Psychological*



- Bulletin*, 107, 65–81.
- Kaplan, S. J. (1983). Cooperative response from a portable natural language system. In M. Brady & R. C. Berwick (Eds.), *Computational models of discourse* (pp. 167–208). Cambridge, MA: MIT Press.
- Kenny, D. A. (1990). Design and analysis issues in dyadic research. *Review of Personality and Social Psychology*, 11, 164–184.
- Kerry, T. (1987). Classroom questions in England. *Questioning Exchange*, 1(1), 32–33.
- King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology*, 14, 366–381.
- King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal*, 27, 664–687.
- Labov, W., & Fanshel, D. (1977). *Therapeutic discourse: Psychotherapy as conversation*. New York: Academic Press.
- Lehnert, W. G. (1978). *The process of question answering*. Hillsdale, NJ: Erlbaum.
- Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, 50, 643–655.
- McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7, 197–244.
- Miyake, N., & Norman, D. A. (1979). To ask a question, one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior*, 18, 357–364.
- Nickel, H., & Fenner, H. J. (1974). Direkte und indirekte lenkung im unterricht in abh ngigkeit von fachspezifischen und methodisch-didaktischen variablen sowie alter und geschlecht des lehrers (Direct and indirect classroom control and its dependence upon subject matter, instructional method, age, and sex of teachers), *Zeitschrift fur Entwicklungs-psychologie und Padagogische Psychologie*, 6(3), 178–191.
- Olson, G. M., Duffy, S. A., & Mack, R. L. (1985). Question asking as a component of text comprehension. In A. C. Graesser & J. B. Black (Eds.), *The psychology of questions* (pp. 219–226). Hillsdale, NJ: Erlbaum.
- Otero, J. C., & Campanario, J. M. (1990). Comprehension evaluation and regulation on learning from science texts. *Journal of Research in Science Teaching*, 27, 447–460.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117–175.
- Papert, S. (1980). *Mindstorms: Children, computers and powerful ideas*. New York: Basic Books.
- Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.
- Pressley, M., Ghatala, E. S., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in response to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly*, 25, 232–249.
- Pressley, M., Goodchild, F., Fleet, J., Zajchowski, R., & Evans, E. (1989). The challenges of classroom strategy instruction. *The Elementary School Journal*, 89, 301–342.
- Pressley, M., & Levin, J. R. (1983). *Cognitive strategy training: Educational applications*. New York: Springer-Verlag.
- Pressley, M., Symons, S., McDaniel, M. A., Synder, B. L., & Turnure, J. E. (1988). Elaborative interrogation facilities in the acquisition of confusing facts. *Journal of Educational Psychology*, 80, 301–342.
- Putnam, R. T. (1987). Structuring and adjusting context for students: A study of live and simulated tutoring of addition. *American Educational Research Journal*, 24, 13–48.
- Reisbeck, C. K. (1988). Are questions just function calls? *Questioning Exchange*, 2, 17–24.
- Resnick, L., Salmon, M. H., & Zeitz, C. M. (1991). The structure of reasoning in conversation. In *The Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 388–393), Hillsdale, NJ: Erlbaum.
- Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.
- Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively*.

- Hillsdale, NJ: Erlbaum.
- Schank, R., Ferguson, W., Birnbaum, L., Barger, J., & Greising, M. (1991). ASK TOM: An experimental interface for video case libraries. In *The Proceedings of the 13th Annual Conference for the Cognitive Science Society* (pp. 570–575). Hillsdale, NJ: Erlbaum.
- Searle, J. R. (1969). *Speech acts*. London: Cambridge University Press.
- Sebrechts, M. M., & Swartz, M. L. (1991). Question-asking as a tool for novice computer skill acquisition. *Proceedings of the 1991 International Conference on Computer-Human Interaction*, 293–297.
- Singer, M., & Donlan, D. (1982). Active comprehension: Problem solving schema with question generation for comprehension of complex short stories. *Reading Research Quarterly*, 17, 166–186.
- Sternberg, R. J. (1987). Questioning and intelligence. *Questioning Exchange*, 1, 11–13.
- Stevens, A., Collins, A., & Goldin, S. E. (1982). Misconceptions in students' understanding. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 13–24). New York: Academic Press.
- Stokal, R. R. (1974). Classification. *Science*, 185, 115–123.
- Susskind, E. (1969). The role of question asking in the elementary school classroom. In F. Kaplan & S. B. Sarason (Eds.), *The psycho-educational clinic*. New Haven, CT: Yale University Press.
- Valdez, G. (1986). Analyzing the demands that courtroom interaction makes upon speakers of ordinary English: Toward the development of a coherent descriptive framework. *Discourse Processes*, 9, 269–303.
- van der Meij, H. (1987). Assumptions of information-seeking questions. *Questioning Exchange*, 1, 111–117.
- van der Meij, H. (1988). Constraints on question asking in classrooms. *Journal of Educational Psychology*, 80, 401–405.
- van Lehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.
- Webber, B. (1988). Question answering. In S. C. Shapiro (Ed.), *Encyclopedia of artificial intelligence: Vol. 2* (pp. 814–822). New York: John Wiley and Sons.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81, 329–339.

Manuscript received June 17, 1992

Revision received December 14, 1992

Accepted April 30, 1993