Running Head: TUTOR LEARNING

Instructional Science (In press)

Tutor learning: The role of explaining and responding to questions

Rod D. Roscoe & Michelene T. H. Chi

Learning Research and Development Center

Department of Psychology

University of Pittsburgh

Correspondence to:

Rod Roscoe

3939 O'Hara Street, Office 561

Learning Research and Development Center

University of Pittsburgh

Pittsburgh, PA 15260 USA

Email: roscoe@pitt.edu

Phone: (412) 624-6488

Fax: (412) 624-7904

Abstract

Previous research on peer tutoring has found that students sometimes benefit academically from tutoring other students. In this study we combined quantitative and qualitative analyses to explore how untrained peer tutors learned via explaining and responding to tutee questions in a non-reciprocal tutoring setting. In support of our hypotheses, we found that tutors learned most effectively when their instructional activities incorporated *reflective knowledge-building* in which they monitored their own understanding, generated inferences to repair misunderstandings, and elaborated upon the source materials. However, tutors seemed to adopt a *knowledge-telling bias* in which they primarily summarized the source materials with little elaboration. Tutors' reflective knowledge-building activities, when they occurred, were more frequently elicited by interactions with their tutee. In particular, when tutees asked questions that contained an inference or required an inferential answer, tutors' responses were more likely to be elaborative and metacognitive. Directions for future research are also discussed.

Keywords: peer tutoring, tutor learning, tutorial dialogue, explanations, questions, metacognition, verbal data analysis

Tutoring is one of the most effective educational interventions, with tutored students performing one to two standard deviations above non-tutored students (Bloom, 1984). Ideally, every student would receive expert tutoring, but this is implausible due to high financial costs and a limited availability of tutors (see Saulny, 2005). Peer tutoring, the use of students as instructors of other students, provides a way to leverage the potential of tutoring that minimizes these limitations (Greenwood et al., 1988). Student tutors are plentiful, and may be recruited as volunteers, compensated with course credit, or hired for a modest wage. More importantly, research has shown that students can be effective tutors, with an average Cohen's d effect size (*ES*) of about .40 (Cohen et al., 1982). Although these gains do not surpass those of expert tutoring, peer tutoring offers an additional unique academic benefit. Student tutors can learn from their teaching experiences (Cohen et al., 1982; Rohrbeck et al., 2003), which we refer to as the *tutor learning effect*. When peer tutoring works, two students benefit simultaneously.

This tutor learning effect has been observed across tutoring formats, such as cross-age (Allen & Feldman, 1976), same-age (Annis, 1983), and reciprocal tutoring (Fantuzzo et al., 1989). Benefits have also been shown in diverse domains, such as reading (Juel, 1996), math (Sharpley et al., 1983), history (Annis, 1983), psychology (Fantuzzo et al., 1989), and biology (Coleman et al., 1997). Encouragingly, tutoring also benefits tutors with low prior achievement (Sharpley et al., 1983) or learning disabilities (Scruggs & Osguthorpe, 1986). However, despite such wide applicability, the average effect size (.33) of tutors' learning gains compared to students learning the same material in non-tutoring settings is not strong (Cohen et al., 1982). Moreover, there can be large variability in reported tutor learning outcomes. In one meta-analysis (Rohrbeck et al., 2003), about 25% of the sampled studies showed strong effects ($ES \geq .80$) for tutors over non-tutors, but 15% of studies showed null results or a disadvantage for tutors. In sum, there is room for improvement in the consistency and magnitude of tutor learning gains. One way to address these issues is to directly examine tutors' actual behaviors, and how these behaviors support or inhibit their learning (Fogarty & Wang, 1982).

How do peer tutors learn by tutoring?

Educators have hypothesized that tutor learning is a direct result of tutors' engagement in instructional activities inherent to the tutoring process, such as explaining, answering questions, correcting tutee errors, manipulating different representations, etc. (Cohen, 1986; Gartner et al., 1971; King, 1998). Such activities are believed to provide opportunities for peer tutors to rehearse their knowledge, and more importantly, to integrate new and prior knowledge and generate new ideas. Tutoring a peer may also allow tutors to monitor their own understanding, and recognize and repair knowledge gaps and misconceptions. Both of these processes may work reciprocally. As tutors share and construct ideas, these ideas become available for self-evaluation and reflection. Ideas that are found to be incorrect or lacking may then be revised or reconstructed. Collectively, we refer to these intermingled processes of knowledge construction and metacognition as *reflective knowledge-building*.

In this project, we focused on the role of reflective knowledge-building in tutors' learning from explaining and responding to questions. We chose these activities because they are fundamental to the tutoring interaction (Graesser, et al., 1995) and have been previously considered as sources of tutor learning (e.g. King et al., 1998). In the following sections, we elaborate upon how explaining and responding to questions may support tutor learning via reflective knowledge-building.

Explaining and reflective knowledge-building

Explanations are an ubiquitous tool used by tutors to communicate key concepts, principles, and relationships, and to correct tutee errors. Such explanations may entail summarizing main ideas, developing analogies, working through examples, etc. Not surprisingly, the richness of this activity has led many to hypothesize that tutors learn by explaining (Bargh & Schul, 1980; Cohen, 1986; Coleman et al., 1997; King, 1998). For example, simply generating and articulating these explanations may improve recall of the material (Osborne & Wittrock, 1983; Pressley et al., 1987), and similar benefits may arise from selecting and summarizing main ideas (e.g. Brown & Day, 1983).

Additional opportunities for knowledge-building may arise from specific explaining acts. One such activity is generating analogies to make abstract concepts more understandable to tutees. For example, in explaining the functions of the retina in the human eye, a peer tutor might compare it to the film in a camera. In creating such analogies, peer tutors must reason through the conceptual mappings, which might facilitate deeper understanding of those relationships (e.g. Duit et al., 2001). Tutors may also use examples to illustrate important concepts and procedures. Research on learning from worked-out examples has shown that learners gain a deeper understanding by decomposing the example steps and linking them to underlying principles (e.g. Atkinson, Renkl, & Merrill, 2003; Chi et al., 1989). Peer tutors may similarly benefit when they carefully break down examples and explain them with their tutees.

The instructional nature of peer tutors' explanations offers additional demands which may encourage reflective knowledge-building. For tutors' explanations to be effective, they need to be relevant, coherent, complete, and accurate (Coleman et al., 1997; Leinhardt, 2001; Webb, 1989; King, 1994). However, although peer tutors may know more than their tutees, they are still students and thus their domain knowledge could be novice-like in a number of ways. Peer tutors may still possess knowledge gaps and misconceptions, and their knowledge may be disorganized. As tutors work to produce effective explanations, they may have to first confront such deficits.

This aspect of learning from explaining draws heavily on the "reflective" component of reflective knowledge-building. Tutors may need to utilize metacognitive processes such as comprehension-monitoring and metamemory, which involve evaluation of the quality of one's own knowledge and understanding (e.g. Flavell, 1979; Hacker, 1998; Maki et al. 2005). As students progress through any learning task, they can judge themselves by various criteria (e.g. subjective feelings of confusion, ability to recall facts on flash cards, etc.) to determine whether they have learned the material. The peer tutoring task may further involve tutors' judgments about how well they are able to explain the information in a way that is understandable and useful.

Based on such judgments, peer tutors might choose to elaborate or revise their understanding. In order to produce relevant explanations that address core topics, tutors may have to first have prioritize the information and decide which concepts are most related. Generating coherent and organized explanations may require tutors to reorganize their own flawed mental models by rearranging conceptual connections. Similarly, striving to create explanations that fully address all concepts may lead tutors to evaluate the breadth and depth of their own knowledge. The need to offer explanations that state main ideas accurately may cause tutors to reevaluate their own comprehension of the material, and test whether their explanations make sense. In short, explaining contains many opportunities for tutors to generate ideas and repair their own knowledge deficits. These processes should result in a more organized, correct, and/or deeper understanding of the material (Bargh & Schul, 1980; Coleman et al., 1997; King et al., 1998).

It is not surprising that reflective knowledge-building processes have been shown to support learning in non-tutoring settings. For example, self-explaining is a strategy in which learners monitor their own comprehension and then generate inferences to fill knowledge gaps and revise misunderstandings (Chi, 2000). Research on self-explaining has found that it contributes to successful problem-solving (Chi et al., 1989, physics domain) and conceptual understanding (Chi et al., 1994, circulatory system). Similar results have been obtained with peer collaboration. For example, King (1994, physiology) and Coleman (1998, photosynthesis) have supported impressive learning gains by training elementary school peers to use prompts encouraging explanation, use of prior knowledge, critical debate about ideas, and comprehension monitoring.

Fewer studies have directly examined peer tutors' explanations and how they contribute to tutor learning. Some researchers have examined how cross-age tutors' behaviors in naturalistic (i.e. no training) settings contribute to *tutee* outcomes. These studies did not collect data on tutor learning, but provided insights about tutors' typical behaviors. Such studies found that explaining was a central activity for middle school math and computer literacy tutors working with elementary school students (Fogarty & Wang, 1982), high school algebra tutors tutoring middle school students and graduate school research methods tutors working with undergraduates (Graesser & Person, 1994; Graesser et al., 1995), and college biology tutors working with other undergraduates (Chi et al., 2001). These studies also showed, however, that these explanations tended to be lengthy and didactic, containing examples and definitions pulled directly from source materials, and were associated mainly with tutees' shallow learning. Tutors rarely generated new elaborations or engaged in deep reasoning discussions, even though these tutoring moves were more likely to lead to deeper tutee learning.

Others have studied tutors' explanations and learning in the context of training studies in which tutors are instructed to use specific tutoring strategies. Such studies speak to the efficacy of the assigned strategies, but may ignore other potentially beneficial activities. For example, Coleman et al. (1997) found that undergraduate tutors trained to teach a same-age peer about evolution by "explaining" (i.e. elaborating with inferences) outperformed tutors told to teach by "summarizing" (i.e. paraphrasing without elaboration). Tutorial dialogues were not directly analyzed, however, and so it is not clear whether tutors in either group actually engaged in reflective knowledge-building or not.

Fuchs and colleagues have examined math tutors' explaining behaviors more directly (e.g. Fuchs et al., 1994; Fuchs et al., 1997). For example, Fuchs et al. (1997) found that elementary-school reciprocal peer tutors trained to give "conceptual explanations" in math (i.e. explain underlying concepts and create examples) outperformed reciprocal tutors who were only told to give feedback when tutees made mistakes. The explanation tutors also produced more statements that integrated the material and asked more problem-solving questions. In stark contrast, the comparison tutors gave "lectures" that merely described concepts without elaboration, and gave tutees few opportunities to participate. However, no data were reported regarding how tutors may have monitored their own comprehension.

In sum, peer tutors who generate explanations that incorporate knowledge-building elaborations seem to learn more effectively than tutors who only summarize (Coleman et al., 1997; Fuchs et al., 1997). However, it is not clear when or why such elaborations occur, and there is no evidence to assess how learning from explaining involves tutors' self-monitoring. Another key idea is that untrained peer tutors seem to adopt a *knowledge-telling bias*. Instead of building upon their knowledge, tutors often "tell what they know" by summarizing and giving didactic lectures (e.g. Chi et al., 2001; Graesser et al., 1995; Fuchs et al., 1994; Fuchs et al., 1997), although untrained tutors can sometimes engage in productive activities spontaneously (e.g. Graesser et al., 1995).

Responding to questions and reflective knowledge-building

Another crucial element of the tutoring process is the interaction that occurs between the tutor and tutee. Tutors do not explain into a vacuum; tutees can ask questions that directly influence the nature of the dialogue that occurs. Particularly interesting are tutees' information-seeking questions, which arise when learners (tutees, in this case) notice a contradiction, or perceive a lack of knowledge, causing them to feel confused or curious (e.g. Graesser & McMahen, 1993; van der Meij, 1994). Such questions differ depending on the type of information needed, such as "shallow" questions about definitions or simple calculations, or "deeper" questions about causal relationships and underlying principles (see Graesser & Person, 1994, for a detailed discussion about the kinds of questions that can occur during tutoring).

From a tutor learning perspective, tutees' questions may be beneficial in several ways. In one case, tutors' knowledge and explanations are accurate and complete, but the tutees' confusion is deep or robust. Or perhaps the tutee is curious about the topic and wants to know more than the tutor initially provides. In this scenario, tutee questions may push the tutor to find new or alternative ways to explain the same ideas. For example, tutors might need to construct a novel example showing how the ideas apply to the real world, or draw a new diagram linking main ideas. Thus, even when tutors know the material fairly well, tutee questions could lead them to further build upon that knowledge.

Alternatively, tutee confusion could arise because peer tutors' knowledge and explanations are lacking, perhaps omitting key ideas or describing a concept incorrectly. In this situation, tutee questions could act as unintentional metacognitive cues that the tutor's knowledge is flawed. Tutors may initially believe that their understanding is accurate, but tutees' confusion engenders a sense of cognitive conflict similar to conflicts between readers' knowledge and text information (e.g. Chan, 2001). This could cause tutors to reexamine their own beliefs. If tutors recognize themselves as the source of the problem, they may reorganize their ideas or generate inferences to repair the errors. In sum, tutee questions may provide further opportunities for peer tutors to engage in reflective knowledge-building (e.g. Bargh & Schul, 1980; Cohen, 1986; Foot et al., 1990; King et al., 1998). Tutee questions may actually stimulate a meaningful proportion of tutors' learning through elaboration and self-monitoring.

Research on answering questions in non-tutoring settings has typically shown this activity to be beneficial to learning. Not surprisingly, these findings also indicate that deeper questions support learning more effectively than shallow questions. For example, well-established research on questions embedded in text have found that questions requiring elaboration, inference generation, and logical reasoning lead to significantly stronger learning than shallow factual questions (e.g. Hamaker, 1986; Peverly & Wood, 2001). Other work has noted benefits for questions embedded in computer-based learning environments (e.g. Azevedo & Hadwin, 2005; Davis, 2003). Davis (2003) inserted prompts such as "Pieces of evidence we didn't understand very well included..." into science learning units. Collaborating middle school peers working with the prompts learned more effectively, and showed more frequent self-monitoring, principle-based explanations, and integration of ideas than unprompted students.

As with explanations, few studies have examined the role of questions in tutor learning. Graesser & Person (1994) found that their tutors and tutees primarily asked shallow questions about definitions and verification of facts, but deep reasoning questions about causal relationships and interpretations of data did occasionally occur. No systematic data were reported regarding how tutors responded to different kinds of tutee questions. However, Chi et al. (2001) have noted that tutee questions could sometimes stimulate interactive dialogues that scaffolded tutees' deeper learning. No tutor learning data were acquired in these studies.

King et al. (1998) examined learning by responding to questions by training middle school reciprocal tutors to use different questioning strategies. All tutors were instructed to give explanations that focused on the "how" and the "why" of various physiological systems (e.g. circulatory system and nervous system). Some tutors were further trained to ask "review" and "thinking" questions that prompted for integration of new and prior knowledge. These questioning groups outperformed the explaining-only groups. Regarding tutors' behaviors, the explaining-only group asked few questions and tended to summarize basic facts in a knowledge-telling manner. The questioning groups produced more explanation statements that integrated and applied the material, and asked deeper questions that prompted critical thinking. These results were recently replicated by Ismail & Alexander (2005), using high school students enrolled in a physics course. However, neither studied directly analyzed whether deeper tutee questions *directly* elicited reflective knowledge-building responses. No data were presented regarding whether tutor learning was accompanied by metacognitive self-monitoring.

In sum, available research offers tentative support for the hypothesis that tutee questions support reflective knowledge-building and tutor learning. When tutees asked deep questions asking for elaboration and integration, tutors seemed to respond with more productive explanations (Chi et al., 2001; King et al., 1998; Ismail & Alexander, 2005). However, no study has shown this link explicitly or has directly considered how different questions may facilitate tutors' metacognitive self-monitoring.

Hypotheses regarding tutor learning behaviors

In the remainder of this article, we further explore several hypotheses that emerge from prior research on tutor learning. One hypothesis is that tutors will learn most effectively when their explanations incorporate reflective knowledge-building (i.e. comprehension-monitoring, integration, and inferences) rather than knowledge-telling (i.e. summarization and paraphrasing). This is not a controversial hypothesis because prior studies have shown that training tutors to use knowledge-building strategies improves learning outcomes (e.g. Fuchs et al., 1997). However, the effort of generating knowledge-telling explanations might still help tutors reinforce their recall of the information (Pressley et al., 1987). Thus, reflective knowledge-building may not be the only route to tutor learning, although we hypothesize that it will be the stronger path.

A more interesting question is whether and how reflective knowledge-building actually occurs. Prior studies have found that less trained tutors produced more shallow and unelaborated explanations (Fuchs et al., 1997; King et al., 1998). Thus, we hypothesize that peer tutors will exhibit a knowledge-telling bias in which their explanations incorporate more knowledge-telling than reflective knowledge-building. However, we also believe that some peer tutors can engage in reflective knowledge-building activities even without training (e.g. Graesser et al., 1995). Two potential contributing factors are considered. First, before tutors revise their explanations, they must perceive a need to do so. Thus, learning and knowledge-building should be linked to metacognitive self-monitoring. Second, research suggests that tutor-tutee interactions, particularly responding to tutee questions, will help to stimulate peer tutor responses that incorporate reflective knowledge-building (Ismail & Alexander, 2005; King et al., 1998). Deeper tutee questions should be more effective prompts than shallow tutee questions.

Method

In this project we focused on untrained peer tutors instructing another student in a fixed-role context (i.e. non-reciprocal). The study also took place in a laboratory setting in which typical programmatic supports for tutor learning were absent. This approach allowed us to be sensitive to the learning opportunities *inherent* to the tutoring process. When tutors are trained to engage in specific strategies, important spontaneous tutoring behaviors that positively or negatively impact tutor learning may be obscured. Similarly, in reciprocal tutoring the participants learn from both tutoring and being tutored, making it difficult to assess the specific contribution of the tutoring role. Many tutoring programs also provide support for tutor learning that goes beyond the actual tutoring experience, such as additional instruction or discussion of the subject matter among the tutors (e.g. Juel, 1996). Our design bypassed these potential confounds.

Our research also utilized quantitative analyses of actual tutorial dialogue data. This method of verbal data analysis allowed us to consider our hypotheses at several levels, but was also time-consuming and difficult (Chi, 1997). Transcription and coding require many months and multiple passes, leading to a tradeoff between statistical power based on sample size and analytical power based on depth of coding. For this partly exploratory study, we chose to work with a smaller sample, aware of the constraints this placed on our statistical power and interpretation. Thus, although we report statistical tests throughout the paper, they should be interpreted cautiously. The overall patterns of results are more telling than the outcomes of individual analyses.

However, it should also be noted prior research using similar analyses and sample sizes have provided rich and informative results. Research on self-explaining has employed careful analyses of "think-aloud" protocols to reveal important patterns in learners' reading and problem-solving behaviors (Chi, 2000). Research on experienced and expert tutors has also greatly benefited from intensive analyses of small numbers of tutors (e.g. Derry & Potts, 1998; McArthur et al., 1990; Merrill et al., 1995; Putnam, 1987; VanLehn et al., 2003). Such research has contributed significantly to our understanding of how expert tutors scaffold tutees' reasoning and problem-solving, diagnose and remediate errors, provide feedback, and adapt to individual tutees.

Participants

Forty college undergraduates, recruited through advertisement flyers posted across the university campus, volunteered to participate in this study for cash payment. Thirty students participated in one of three experimental conditions that manipulated different situations for tutoring and explaining. Ten students were recruited as tutees. To ensure that participants had minimal background knowledge about the domain, students who had taken certain biology, physiology, or neuroscience courses were not eligible to participate. A summary of demographic information is presented in Table 1. Group comparisons (ANOVA) of mean age, verbal SAT scores, and GPA, and chi-square tests of gender, year in school, and academic major, showed that participants were fairly evenly distributed across conditions on all demographic factors.

-------- Insert Table 1 about here --------

Materials

Human eye and retina text. The domain selected was the basic structure, location, and function of human eye and retina components. Some prior studies on tutor learning (e.g. King et al., 1998) and explaining (e.g. Chi et al., 1994) have utilized a physiology domain, and so our choice of domain offers comparability to such work. This information was presented in a short text discussing the regulation and focusing of light via the cornea, iris, lens, and related structures (see Appendix). The text also described the function and distribution of photoreceptors in the retina, and the interconnections between several retinal neuron types (e.g. bipolar and ganglion cells). Although the text indicated that neural signals travel to the brain via the optic nerve, it did not discuss any of the visual processing that occurs in the brain. Two labeled diagrams were provided: a cross-section drawing of the entire eyeball and an illustration of a magnified portion of the retina. Thus, the text included both familiar, everyday topics (e.g. the iris and pupil) and unfamiliar, technical concepts (e.g. neural connections of bipolar cells), providing ample opportunities to make connections with prior knowledge and explore new ideas.

This material was adapted from a college textbook on sensation and perception (Goldstein, 1999) in consultation with human physiology textbooks (e.g. Fox, 1996). For our purposes, the information was reduced to the level of basic definitions of the components and short descriptions of their functions. All examples and analogies were removed, as were advanced references to biochemistry (e.g. synaptic transmission), optics (e.g. inversion of light by the lens), and disorders of the eye (e.g. glaucoma). The resulting text was 68 sentences in length (1025 words), divided into 17 topic-based paragraphs presented on separate pages without headings. Readability was assessed using tools available through Coh-Metrix (McNamara et al., 2005), which indicated a Flesch Reading Ease score of 56 and a Flesch-Kincaid Grade Level score of 9.2.

Knowledge assessments. Learning outcomes were assessed using two written measures (see Appendix). The first of these was the Definitions Test, in which students were instructed to label blank versions of the eye and retina diagrams (target components were indicated by arrows) and provide definitions for each labeled component. Students were encouraged to be accurate and thorough. Participants were given a list of potential terms, some of which were filler items irrelevant to the eye or not pictured on the diagrams. There were 13 targets and 13 filler items on the eye diagram, and 7 targets and 9 filler items on the retina diagram. This test measured recall of factual knowledge about the eye and can be viewed as a measure of shallow learning.

The second measure was the Questions Test, in which students responded to short-answer questions blending recall, integration, and application of the material (adapted from Roy, 2001). For example, one question asked about likely causes of colorblindness. To answer correctly, participants had to a) recall that cone receptors are sensitive to color, b) infer that colorblindness could result from a defect in the cones, and c) infer that cones may be color-specific. Such questions required participants to integrate text contents and generate new knowledge. Participants were instructed to answer the questions as completely and accurately as possible, and were also informed that some questions might not be answerable just memorizing the text information. The questions were selected to address most of the text concepts at least once, with some central concepts (e.g. receptors and cornea/lens functions) tested in more than one question. This test can be considered a measure of participants' deeper learning.

There were two versions of this test: a shorter version with 12 questions given after the initial reading of the text, and a longer version consisting of 20 questions given as a final post-test. The post-test included the original 12 "repeat" questions along with 8 additional "new" questions. Some new questions were randomly mixed with the previous questions, but most were clustered at the end. The purpose of the new questions was to test learners' ability to answer questions they had not previously seen.

Procedures and Conditions

The study was divided into two sessions, one week apart, to facilitate recruitment. In the first session, all participants and tutees took the Definitions Test as a "pre-test" measure of incoming knowledge. Experimental participants (not tutees) then studied the visual system text for about 30 minutes. At this point, participants were not informed of the nature of the second-session task, but only that they would be reviewing the material again using some assigned strategy. The purpose of this design was to 1) allow tutors the opportunity to develop the necessary background knowledge of the subject matter before tutoring while maintaining equal exposure for all participants, and 2) eliminate outcome differences that might arise due to some participants studying with a “teaching expectancy.” Research on tutor learning has found that students who studied new material with the expectation of later teaching it learned the material better than students who studied with the expectation of only taking a test (e.g. Annis, 1983; Bargh & Schul, 1980; Benware & Deci, 1984; although see Renkl, 1995). In this study, we wished to understand how tutors learn from the tutoring process itself, not from preparing-to-teach. Our design allowed us to reduce the possibility that task expectations affected the way participants initially studied the material. At the end of the first session, experimental participants took the Definitions Test and the short version of the Questions Test as “post-reading” measures of what they learned from reading the text.

In the second session, participants were randomly assigned to one of three experimental conditions described below (tutees were recruited as needed for the tutoring condition): peer tutoring, tutorial explaining, or self-explaining. Participants were given 30 minutes for this activity and were videotaped as they worked. Afterwards, all participants and tutees took the Definitions Test and the long version of the Questions Test as “post-test” measures of what they learned from their respective tasks.

Peer tutoring. In this group, ten undergraduate students (peer tutors) explained about the human eye to another student (tutees). Tutors were told to explain by “going beyond what the text says” and to “help the listener to really understand,” but received no training. Tutors were also encouraged to answer the tutees' questions and had the text available for reference. Tutees were instructed to learn by paying attention and asking questions, but were never allowed to read the text. The tutees took the Definitions Test and Questions Test only as pre-test and post-test assessments. This condition captured a fully interactive setting for same-age tutoring and explaining, meaning that both the tutor and tutee could contribute to the tutorial dialogue. This interaction was asymmetrical, however, because tutors and tutees did not have equal knowledge or access to materials.

Tutorial explaining. In this group, ten undergraduates (tutorial explainers) produced a videotaped explanatory lesson about the eye. These explainers were instructed to produce a lesson that could be used by another student, at about their same level of education, to successfully learn the material. As with the peer tutors, no formal training was provided, but explainers were encouraged to go beyond the contents of the text and explain in a way that would help another student to learn. This condition captured a non-interactive setting for explaining, in which tutorial explainers could engage in whatever instructional explaining activities they desired, but without any feedback or questions from a tutee. A comparison of the peer tutors and tutorial explainers may provide insight into some of the effects of tutor-tutee interactions on tutor learning and explaining.

Self-explaining. In this group, ten undergraduates (self-explainers) reviewed the text while explaining aloud to themselves. Similar to the other conditions, self-explainers were encouraged to "go beyond what the text says" and "try to really understand" the material. Self-explainers were told to talk aloud the entire time, but no one was present to prompt them to talk. This condition provided a contrasting learning-by-explaining task that has been previously shown to support knowledge-building, but which involved no instructional demands. Comparisons to each of the other conditions may further reveal how tutoring tasks and tutor-tutee interactions influence learning from explaining.

Analyses and Results: Knowledge Assessments

Scoring of the Knowledge Assessments

The Definitions Test was scored by tabulating the number of correct definitions generated. Correct definitions were defined as any piece of accurate and relevant information about a given eye component. This measure had no absolute maximum score but the highest score obtained on the post-test was 59 and the lowest was 14. The Questions Test was scored by tabulating the number of correct responses students gave in their short answers. Correct responses were defined as any accurate and relevant statement that answered the core question being asked (i.e. irrelevant information was ignored), thus allowing participants to earn “partial credit” for incomplete responses. This measure had no absolute maximum score but the highest score obtained on the post-test was 43 and the lowest was 7. There was some subjectivity in deciding whether responses were correct, although the text and other resources (e.g. textbooks) were used to assess the accuracy of answers. Coding reliability was established by having two raters (first author and a research assistant) score the same randomly selected set of 15 subjects’ data. There was 94% agreement, and disagreements were resolved through discussion.

Performance of the Tutees

Although our focus is on tutor learning, we briefly discuss tutee performance to show that the tutors were successful instructors. On the Definitions Test, tutees generated an average of 5.8 correct definitions at pre-test and 24 correct definitions at post-test, a significant gain, $t(9) = 7.4$, $p < .001$. On the Questions Test, tutees produced 5.5 correct responses at pre-test and 24.1 correct responses at post-test, also a significant gain, $t(9) = 7.2$, $p < .001$. Such tutee learning gains are rather impressive considering their tutors were untrained undergraduates, who had only initially studied the material for 30 minutes a week prior to the tutoring sessions. The peer tutors were fairly effective instructors.

Effects of Task and Tutor-Tutee Interaction on Learning Outcomes

Table 2 displays mean knowledge assessment scores at each time point for each condition. All analyses throughout the paper were conducted as analyses of variance (ANOVA) unless otherwise noted, followed by Bonferroni pair-wise comparisons. We chose a standard alpha level of .05 for all tests, although statistical significance is sometimes of less importance than the overall patterns of effects observed considering our smaller sample size. Effect sizes (Cohen's d) for key comparisons are provided.

-------- Insert Table 2 about here --------

No significant differences were found on the pre-test or post-reading Definitions Test, suggesting that all groups entered the study with similar (and minimal) background knowledge and acquired comparable factual knowledge from their initial reading of the text. At post-test, a significant group difference was obtained, $F(2,27) = 8.30$, $p < .01$. Pair-wise comparisons revealed that peer tutors and self-explainers were equivalent at post-test, and both scored significantly higher than tutorial explainers. Thus, self-explainers and peer tutors were able to generate more correct definitions of eye and retina components than tutorial explainers.

On the post-reading Questions Test, a significant group difference was observed, $F(2,27) = 4.78$, $p < .05$. The peer tutors and self-explainers were equivalent and both scored higher than the tutorial explainers, indicating that these two groups understood the material better than tutorial explainers after the initial reading phase. Controlling for this initial difference using ANCOVA procedures, a significant difference was still found between groups at post-test, $F(2,26) = 6.76$, $p < .01$, with self-explainers performing better than peer tutors and tutorial explainers. Although the difference between the peer tutors and tutorial explainers was not significant, it followed the same trend as observed on the Definitions Test.

In sum, self-explainers seemed to gain a deeper understanding of the material than peer tutors and tutorial explainers, and peer tutors appeared to understand the material somewhat better than tutorial explainers. This pattern is further elaborated by breaking down the Questions Test post-test results based on “repeat” and “new” questions. There was an overall effect for both repeat questions, $F(2, 26) = 6.43$, $p < .01$, and new questions, $F(2,26) = 3.61$, $p< .05$. Although pair-wise comparisons did not show statistically significant differences, the trends are informative. For repeat questions, self-explainers scored somewhat higher than peer tutors, who scored slightly higher than tutorial explainers. Interestingly, self-explainers seemed to achieve a higher mean score on the new questions than both peer tutors and tutorial explainers.

What do these patterns tell us about the benefits of tutor-tutee interactions for tutor learning and explaining? First, opportunities to interact with the tutee seemed to support learning for tutors. Students who tutored a peer in an interactive setting appeared to be better able to define key terms and answer comprehension questions than students who generated tutorial explanations without tutee interaction. This suggests that some aspect of the tutor-tutee interaction directly or indirectly influenced tutors’ learning outcomes or activities. However, the most successful learners in this study were the self-explainers. Although self-explaining and peer tutoring were equally effective for learning basic facts, self-explaining seemed to better facilitate learning with deeper understanding. An interpretation of this result is that participants may have benefited more by explaining to themselves in a “learner” role than explaining to another person in an “tutor” role, an idea we will return to in the conclusion of the paper.

Analyses and Results: Explaining Protocols

Was the overall pattern of learning outcomes and test scores (i.e. self-explaining > peer tutoring > tutorial explaining) also reflected in participants’ behaviors? We have predicted that learning should be linked to the ways in which participants monitor and build upon their understanding while explaining. Participants who showed stronger learning should be those who engaged in more reflective knowledge-building.

Several analyses were conducted to test hypotheses regarding effective tutoring and learning activities and the influence of tutor-tutee interactions. Specifically, we analyzed the students’ verbal behaviors as they generated and revised explanations in their respective conditions. We did not examine other types of behaviors, such as drawing or gesturing. In the following sections, we first describe episodes of reflective knowledge-building and knowledge-telling that occurred as students explained, and test relationships between these activities and learning. We then focus on the peer tutors to assess whether tutors or tutees were responsible for initiating these activities, and the impact of tutee questions on tutors’ reflective knowledge-building.

Coding of Knowledge-Telling and Knowledge-Building

All of the protocols were fully transcribed and segmented according to changes in the topic of discussion, which formed the boundaries of explanation “episodes.” Thus, each episode was a brief segment of the overall explanation that was devoted to one particular topic. These topics could refer to a single eye component (e.g. rod receptors) or interactions between multiple components (e.g. how rod and cone receptors are involved day and night vision). The episodes were then characterized by the overall pattern of statements uttered. For example, paraphrase statements simply reiterated text contents verbatim, or with some rewording or clarification. In contrast, inference statements contained information that was not given in the text, and thus had to be integrated from prior knowledge or newly generated. Metacognitive statements contained a self-evaluation of the speaker's comprehension, accuracy, uncertainty, etc.

Overall, 996 codeable episodes were identified in this manner. About 33 episodes were produced per participant on average, and the overall number of episodes produced was significantly correlated with Definitions Test scores, $r(28) = .50$, $p < .01$, but not with Questions Test scores, $r(28) = .17$, *ns*. Thus, students who covered more topics gained better factual knowledge of the domain, which is not surprising. However, simply explaining more did not guarantee deeper understanding. There was also some variation across conditions, $F(2,27) = 4.84$, $p < .05$. Peer tutors produced the most episodes on average and tutorial explainers produced the fewest episodes. Only the difference between peer tutors and tutorial explainers was statistically significant. Although it is tempting to conclude that peer tutoring encouraged more explanation than other settings, this result may be partially artifactual. The peer tutoring condition involved two speakers who interrupted each other and thus could shift topics more rapidly. In regards to our hypotheses, it is much more important to consider the actual nature of these episodes.

Coding of episodes. We initially coded each episode as an instance of one of two broad classes of activities, knowledge-telling or knowledge-building, as defined by our theoretical framework. Each episode was also coded as to whether it contained overt self-monitoring statements. Several subcategories of knowledge-telling or knowledge-building activities emerged during this process, which were then incorporated into the complete coding scheme. Reliability of the episode coding process was established by having two coders (first author and a graduate student) code the same randomly selected subset of six complete protocols (2 from each condition). The second coder was presented with a template describing the features of each episode type, along with examples, which were explained verbally. Agreement between coders was high (96%, kappa = .90), and disagreements were resolved through discussion.

Knowledge-telling. Episodes coded as “knowledge-telling” consisted primarily of paraphrase statements. Most knowledge-telling episodes were unelaborated summaries of the text information, and some were little more than definitions of eye components pulled directly from the text. In some cases, students also created shallow mnemonic devices based on rhyming words or matching letters. Not surprisingly, the frequency of knowledge-telling episodes was only moderately correlated ($N = 30$) with post-test Definitions Test scores; $r(28) = .25$, and somewhat negatively correlated with Questions Test scores; $r(28) = -.12$. Neither correlation was statistically significant (Table 3).

Excerpt A, generated by a tutorial explainer, shows a typical knowledge-telling episode. In this example, the explainer summarizes a text passage (see Appendix) describing the connections of the horizontal and amacrine cells. Many of the explainer’s statements are nearly verbatim paraphrases of the text.

(A) There are two other types of main retinal cells (paraphrase) called horizontal cells and amacrine cells. (paraphrase)

Which I can show after the text. (comment to the viewer)

The main function of the horizontal cells is to connect receptors to other receptors. (paraphrase)

So that's pretty easy to remember. (metacognitive)

Horizontal cells connect receptors to other receptors. (paraphrase)

And the amacrine cells connect ganglion cells to other ganglion cells (paraphrase)

and bipolar cells to other bipolar cells. (paraphrase)

A few knowledge-telling episodes, however, did contain elaborations in which implicit concepts were discussed more explicitly. While summarizing the text, learners sometimes clarified technical terms by putting them into everyday language. This "concept elaboration" activity was considered knowledge-telling because the explainer was sharing information they knew prior to the study. The frequency of concept elaboration was significantly and positively correlated with post-test Definitions Test scores, $r(28) = .38$, $p < .05$, but not linked to Questions Test performance, $r(28) = .15$, *ns*. Thus, the inclusion of even minor elaborations supported improved fact recall.

Excerpt B shows an example of concept elaboration in which a tutorial explainer clarifies the text statement "the cornea must be transparent and free from scar tissue in order for a sharp image to be formed on the retina."

(B) The next most important thing in your eye is the cornea. (paraphrase)

And when people say they scratch their cornea, they damage their cornea (clarify, paraphrase)

And that's bad, because the cornea is an important part. (paraphrase)

It's responsible for 70% of the focusing power of the eye. (paraphrase)

But if it's damaged you can- it damages your sight (clarify, paraphrase)

because scarring on the cornea makes it- it's really bad. (paraphrase)

Because your sight is affected (paraphrase)

It affects your eye's ability to focus (clarify, inference)

The cornea is actually for bending the light that passes through the eye. (paraphrase)

Knowledge-building. Episodes identified as "knowledge-building" involved integration of concepts and generation of knowledge through inferences. In these episodes, students built upon the text information by drawing upon their prior knowledge and logical reasoning to develop a more complete or accurate understanding. Inference statements were commonplace in these episodes, as were metacognitive statements about the learners own knowledge and comprehension. As expected, frequency of knowledge-building episodes was significantly and positively correlated with post-test Definitions Test, $r(28) = .58$, $p < .01$, and Questions Test scores, $r(28) = .47$, $p < .01$ (Table 3).

This knowledge-building activity is exemplified by Excerpt C, in which a self-explainer struggles to deduce the relationship between the changing thickness of the lens and the focusing power of the eye described in the text (see Appendix). Though the student does not produce a perfect final answer, her mental effort is impressive.

(C) So I guess when you have a- so if you contract then that's when I said it would be thinner. (inference)

So I guess when you have a thinner- but why would that be? (metacognitive, self-question)

Because when you have glasses, the stronger your prescription, the thicker your lens is. (inference, integration of prior knowledge)

Unless it's like the opposite like- than from your eye or something. (inference)

But anyways. Just when they get tighter, the refraction power gets higher. (paraphrase)

So I guess as it gets thinner the- I wonder. (metacognitive)

If it tightens it has to- they have to get closer together, (inference)

so it has to be thinner. (inference)

Because otherwise they'd just be spread out. (inference)

Two specific knowledge-building subcategories emerged from our coding of the episodes: elaborated reviewing and sense-making. First, learners often reviewed the material, revisiting concepts for further consideration. While reviewing, some students made additional new connections with prior knowledge, or generated novel examples and analogies to further develop the meaning of text concepts. Thus, instead of simply repeating the same ideas again, these episodes involved rethinking or application of the ideas. Frequency of these "elaborated reviewing" episodes was positively and significantly correlated with both the post-test Definitions Test, $r(28) = .50$, $p < .01$, and Questions Test, $r(28) = .49$, $p < .01$.

Excerpt D shows an example of elaborated reviewing in which a peer tutor reviews the functions of the iris muscles. He freely admits that he did not fully understand the topic previously. In reviewing the material, he is able to revise his understanding of the iris muscles' actions and functions.

(D) Alright. This is something that I didn't really get before. (metacognitive)

When the circular muscles contract, they're squeezing the iris closed (inference)

and the pupil gets smaller. (paraphrase)

When the radial muscles contract, they're pulling back, away, from the center. (inference)

So they're pulling it open. (paraphrase)

So that's why you need two different kinds of muscles. (inference)

I never really got that before. That's pretty cool. (metacognitive)

So it isn't the same muscle going back and forth. (inference)

It's two different muscles for opening and closing. (inference)

In "sense-making" episodes, learners drew upon their prior knowledge and generated new inferences in order to satisfy their own curiosity, or to repair perceived errors and confusions (similar to self-explaining). Such episodes were characterized by frequent uncertainty and guesswork; learners were often working at the boundaries of their knowledge. As with elaborated reviewing, the frequency of sense-making was significantly and positively associated with both post-test Definitions Test scores, $r(28) = .37$, $p < .05$, and Question Test scores, $r(28) = .38$, $p < .05$.

Excerpt E below (see also Excerpt C) shows an example of sense-making. In this excerpt, a peer tutor discusses the relationship between dark vision, rod receptors, and receptor placement. What is most interesting is the willingness of the tutor to speculate about novel scenarios (e.g. what if there were no cones in the fovea). This activity was also partly stimulated by a tutee question.

(E) Tutee: Does it depend on the number of rods you have, or--? (deep question)

Tutor: Um. You know what. I don't think it would depend so much on the number. (metacognitive, inference)

It probably does to a degree. (inference)

You do have some rods on the fovea. (paraphrase)

You know, which in our eye the rods begin at least 16 degree out on the fovea (paraphrase)

There are none at all in the center. (paraphrase)

If you do have some in the center, I would imagine you'd be able to resolve fine detail in the dark. (inference, metacognitive)

And say we didn't have any cones in- in the center of the fovea (inference)

We wouldn't be able to see well. (inference)

Despite all the light your wouldn't be able to make out the details. (inference)

Metacognitive monitoring. In addition to identifying whether episodes were examples of knowledge-telling or knowledge-building, episodes were also coded for the presence of metacognitive monitoring. As can be seen in the excerpts above, some episodes contained overt self-monitoring statements. Episodes were classified as either "without monitoring" if no monitoring statements were identified in the episode or as "with monitoring" if one or more metacognitive statements were generated. Although participants likely engaged in some covert metacognitive activity, the occurrence of overt self-monitoring statements was our most direct evidence that monitoring had occurred. Overall, the frequency of episodes with monitoring was positively and significantly correlated with post-test Definitions Test, $r(28) = .58$, $p < .01$, and Questions Test, $r(28) = .42$, $p < .05$. In contrast, episodes that did not contain overt metacognitive statements were only weakly linked to post-test Definitions Test scores, $r(28) = .13$, *ns*, and negatively associated with Questions Test scores, $r(28) = -.22$, *ns*.

It is also important to consider whether metacognitive activity was more likely to occur in knowledge-building episodes, as our theoretical framework would suggest. To test this idea, we considered mean percentages of metacognitive knowledge-building and knowledge-telling episodes across subjects. A paired-samples t-test showed that across all participants ($N = 30$), the mean percentage of knowledge-building episodes with monitoring ($M = 46.5$, $SD = 30.9$) was significantly higher than the mean percentage of knowledge-telling episodes with monitoring ($M = 34.7$, $SD = 17.4$), $t(29) = 2.52$, $p < .05$. Knowledge-building episodes were more likely to be metacognitive in nature.

Patterns of Knowledge-Telling and Knowledge-Building

Our coding of the protocols captured two distinct categories students' explaining activities, knowledge-telling and knowledge-building. The amount of knowledge-building (i.e. generation and integration of knowledge) students engaged in was significantly and positively correlated with both the Definitions Test and Questions Test. However, students' amount of knowledge-telling (i.e. paraphrasing of facts) was not strongly associated with test scores. Importantly, the amount of metacognitive self-monitoring that occurred was also correlated with learning outcomes, and knowledge-building episodes were more likely to be metacognitive than knowledge-telling episodes.

These results highlight an important dissociation. Peer tutors and other learners did not necessarily benefit from simply producing *more* explanations, but from generating more of certain *kinds* of explanations. This pattern is further demonstrated by examining proportion data. The mean proportion of knowledge-building was correlated with the Definitions Test, $r(28) = .44$, $p < .05$, and Questions Test, $r(28) = .46$, $p < .05$, whereas the proportion of knowledge-telling was negatively correlated with test scores (i.e. the same values with the signs reversed, due to the binary nature of the coding scheme). Similarly, the mean proportion of episodes with monitoring was correlated with the Definitions Test, $r(28) = .41$, $p <.05$, and Questions Test, $r(28) = .46$, $p < .05$, where the proportion of episodes without monitoring was negatively correlated with scores.

-------- Insert Table 3 about here --------

Based on the above patterns, differences in the amount of reflective knowledge-building exhibited across groups should parallel observed differences in performance. Self-explainers should engage in more reflective knowledge-building than peer tutors, with tutorial explainers using these strategies least often. To test this prediction, mean frequencies episodes were compared between conditions, which are reported in Table 4 along with effect sizes for keys pair-wise comparisons.

-------- Insert Table 4 about here --------

There was a marginally significant group effect for knowledge-telling episodes, $F(2,27) = 3.17$, $p < .06$. Pair-wise comparisons did not resolve into significant differences between conditions, but peer tutors seemed to engage in more knowledge-telling than both tutorial explainers and self-explainers. Significant group differences were observed for knowledge-building episodes, $F(2,27) = 9.86$, $p < .01$, and episodes with monitoring, $F(2,27) = 12.1$, $p < .01$. Self-explainers and peer tutors engaged in similar amounts of knowledge-building and self-monitoring, and both groups engaged in these activities more often than tutorial explainers.

To further probe these results, we also analyzed the mean proportion of knowledge-telling episodes across conditions (Table 5). This analysis showed a significant effect for group, $F(2,27) = 11.52$, $p < .001$. Tutorial explainers and peer tutors were comparable to each other, and both produced a significantly higher proportion of knowledge-telling episodes than self-explainers. Analyses of the proportion of episodes without monitoring also showed a significant group effect, $F(2,27) = 5.58$, $p < .01$. Although tutorial explainers produced a higher proportion of episodes without monitoring than other conditions, pair-wise comparisons were not significant in this case.

Finally, we considered the three specific strategies identified in the broader categories. There were no main effects for concept elaboration. As one might expect, group differences were observed for elaborated reviewing, $F(2,27) = 8.58$, $p < .01$, and sense-making, $F(2,27) = 5.69$, $p < .01$. Tutorial explainers engaged in these strategies with an astonishingly low frequency – less than one episode per person on average. Pair-wise comparisons suggest that self-explainers and peer tutors were comparable to each other, and generated more of these productive episodes than tutorial explainers.

Summary. This pattern of correlations and group differences indicates that reflective knowledge-building activities were linked to better learning outcomes than knowledge-telling, and distinguished higher scoring conditions from lower scoring conditions. Reflective knowledge-building episodes were positively and significantly correlated with measures of both factual learning and deeper understanding, whereas knowledge-telling episodes were mainly associated with fact acquisition. Moreover, reflective knowledge-building was most often displayed by self-explainers, who outperformed other experimental conditions on knowledge assessments. Peer tutors and tutorial explainers showed a knowledge-telling bias, exhibiting a high frequency and proportion of these less productive knowledge-telling episodes. Interactions with the tutee seemed to partly mitigate peer tutors' knowledge-telling bias. The peer tutors displayed a higher frequency of knowledge-building and self-monitoring than tutorial explainers, resulting in better fact recall and understanding of the material.

Influence of tutor-tutee interaction on peer tutors' behaviors

The preceding analyses suggest that interactions with a tutee supported peer tutors' reflective knowledge-building and learning. In this section, we explore this finding in more depth by focusing on the peer tutoring condition. Two analyses are presented, which first considered who initiated different episodes, and then examined how tutors responded to deep and shallow tutee questions.

Tutor and tutee-initiation of episodes. In the peer tutoring condition, both the tutor and tutee could contribute to the flow of dialogue. As a consequence, some of the tutors' instructional behaviors were self-initiated activities they chose to engage in of their own accord, whereas others were elicited by some action of the tutee. To capture this distinction, episodes were further coded as being either "tutee-initiated" or "tutor-initiated." An episode was coded as tutee-initiated if the tutee selected the concept to discuss, asked a question, or made some comment causing a new episode to begin (i.e. a change in topic). All other episodes were categorized by default as tutor-initiated. Reliability of this coding was established by having two raters (first author and a research associate) code the same randomly selected subset of five complete peer tutoring protocols. There was 96.4% agreement between coders (kappa = .93).

To analyze the pattern of how knowledge-telling and knowledge-building episodes were initiated, we compared the mean percentages of tutor-initiated and tutee-initiated episodes of each type across the ten pairs. Some studies have shown that tutors tend to dominate the tutoring sessions, controlling the selection of topics and how the material is covered (e.g. Graesser et al., 1995). It would not be surprising to find that peer tutors initiated the majority of both episode types. However, we found that 73.9% (SD = 19.4) of knowledge-telling episodes were tutor-initiated on average, compared to 26.2% (SD = 19.4) initiated by tutees, $t(9) = 3.90$, $p < .01$. In contrast, 62.1% (SD = 19.0) of knowledge-building episodes were tutee-initiated on average, whereas 37.9% (SD = 19.0) were tutor-initiated, $t(9) = 2.02$, $p = .075$. These patterns further highlight peer tutors' preference for knowledge-telling as their main style of instruction. Tutees also elicited a majority of tutors' reflective knowledge-building activity.

Tutee questions and tutor responses. Tutee questions could provide a prompt for peer tutors to give more elaborated explanations or reflect on their own understanding. Thus, tutee questions could be one specific way in which tutees might elicit tutors' knowledge-building. Some prior work has found that tutors trained to engage in specific questioning strategies generated better explanations and learned more effectively (e.g. King et al., 1998), but did not explicitly assess the link between tutee questions and tutor responses. In this analysis, we considered the impact of tutee questions on tutors' subsequent knowledge-building and metacognitive monitoring directly.

The episodes identified in previous analyses could contain multiple tutee questions, and so we resegmented the tutoring protocols using "question-response exchanges" as the unit of analysis. A tutee "question" was defined as an interrogative statement in which the tutee requested information or verification of information. We excluded questions about task procedures or off-topic issues. These tutee questions were then coded as "shallow" or "deep." A deep question was one that asked about content not provided explicitly in the text. Thus, these questions either contained an inference that the tutor had to evaluate, or required the tutor to generate an inference in order to answer the question. A tutor "response" was defined as any information or feedback provided by the tutor in answer to a question. Tutor responses were coded as shallow or deep depending on whether they contained inferences or novel references to prior knowledge. Tutor responses were further classified as "metacognitive" or "non-metacognitive" based on whether they contained overt self-monitoring statements.

Reliability of the coding process was established by having two raters (first author and a research associate) code the same randomly selected subset of three protocols (about 110 questions). The second rater was given a sheet showing coded examples of questions and responses, which were discussed verbally. Coder agreement was 87% for question depth (kappa = .75), 91.7% for response depth (kappa = .85), and 97.2% for response metacognition (kappa = .95). Disagreements were resolved through discussion.

Out of a total of 310 questions, 37% were classified as deep and 63% as shallow. Thus, our sample was comparable to previous studies (e.g. Graesser & Person, 1994) showing that shallow questions are more common. We examined the mean percentages of tutee question and tutor response types across the ten tutoring pairs. This analysis showed that, on average, 71.8% (*SD* = 23.8) of deep questions received a deep response, whereas only 19.6% (*SD* = 18.6) of shallow questions received a deep response, $t(9) = 7.24$, $p < .001$. Similarly, 58.1% (*SD* = 14.1) of deep questions elicited a metacognitive response on average, whereas only 27.3% (*SD* = 20.5) of shallow questions elicited responses contained self-monitoring, $t(9) = 4.55$, $p < .01$. Thus, tutees' questions had a meaningful impact on tutors' subsequent elaboration and self-monitoring. Importantly, deep tutee questions were more likely to elicit reflective knowledge-building responses from tutors than shallow tutee questions.

Two excerpts illustrate these question-response relationships. In excerpt D, the tutor initially summarizes information about the retina blind spot and addresses a shallow question. The tutee then asks a deep question that elicits a new analogy from the tutor.

(D) Tutor: This is the blind spot. You can't see anything there because that's where the optic nerve leaves the eye. So there aren't receptors right there. (text paraphrase)

Tutee: Okay, wait. The blind spot is where all the nerves are located? (shallow question)

Tutor: Yeah. Like, that's where all of the optic nerves come together. They go around and that's where they all pull together and go back to the eye. Or back to the brain. So right there, there aren't any receptors. (shallow response, text paraphrase)

Tutee: So how does that affect your vision? (deep question)

Tutor: If something comes in and your lens refracts it to that point then you don't see it. So, it's just like when you're driving and there's that little spot in the mirror where you just won't see the person behind you. It's like that, except for the eyes. (deep response, new analogy)

In excerpt E, a tutor and tutee discuss the iris and pupil. The tutor responds to a shallow question by skimming and paraphrasing the text. However, the tutee's deep question prompts the tutor to reconsider a concept that had confused her, drawing on inferences and prior knowledge to visualize and better understand these eye components.

(E) Tutor: The iris is the colored part of your eye. And it can expand or contract radially or circularly. (text paraphrase)

Tutee: What's radially? Like outward? (shallow question)

Tutor: Um. It explains that on the next page. [skims text] Yeah. That's outward. And when the radial muscles contract, the pupil gets larger. (shallow response, text paraphrase)

Tutee: Okay. So, pretty much… contract is to make it smaller. So wouldn't the iris get smaller? (deep question)

Tutor: Oh! That makes so much sense now. Yeah. Like when your iris gets smaller, your pupil gets bigger. Like when someone's coming out of a dark room or they get surprised. Your pupil gets really big and your iris gets really small. (deep response, new visualization)

Summary. These two focused analyses demonstrate the impact of tutor-tutee interactions and tutee questions on tutors' learning behaviors. Peer tutors demonstrated a clear knowledge-telling bias, initiating the major percentage of knowledge-telling episodes. However, tutees helped to overcome this bias by directly and indirectly eliciting peer tutors' reflective knowledge-building. Most of peer tutors' knowledge-building episodes were initiated by tutees. At a more fine-grained level, our results show that tutee questions were an important part of this process. When tutees asked shallow questions, tutor responses were shallow and non-metacognitive. However, deeper tutee questions that contained or required inferences helped to elicit reflective knowledge-building.

Discussion

Prior research has demonstrated that students can benefit academically from tutoring other students. This tutor learning effect, along with a greater pool of potential tutors and lower cost of implementation, can make peer tutoring a desirable alternative to expert tutoring. However, although peer tutors seem able to learn in a variety of settings, the overall magnitude of the effect could be stronger and more consistent. Some studies show large benefits while others do not (Cohen et al., 1982; Rohrbeck et al., 2003). In this project, we sought to contribute to the small but important body of research that directly examines the processes of tutor learning – how tutors' observable behaviors contribute to learning outcomes. Such research can teach us about how tutors learn, and about what tutoring activities should be promoted or discouraged in order to optimize tutor (and tutee) learning outcomes (Fogarty & Wang, 1982).

One of the major assumptions underlying tutor learning research is that tutoring contains inherent learning opportunities (Bargh & Schul, 1980; Cohen, 1986; Gartner et al., 1971). Tutors are believed to benefit by engaging in fundamental tutoring behaviors, which are hypothesized to support reflective knowledge-building. Trying to produce effective explanations may help tutors monitor their own understanding, and then take steps to repair perceived misconceptions. Tutee questions may act as metacognitive prompts for tutors to generate better explanations or further revise their knowledge.

Prior research on these ideas have focused on training peer tutors to use tactics that are carefully designed to support knowledge-building (e.g. Coleman et al., 1997; Fuchs et al., 1997; Ismail & Alexander, 2005; King et al., 1998). Comparisons were then made between trained and untrained tutors. Not surprisingly, trained tutors showed stronger learning outcomes than untrained tutors, and also engaged in target behaviors more often. An implicit conclusion of this work is that effective tutor learning strategies may have to be externally inserted into the tutoring process. In this study, we showed that untrained tutors could spontaneously engage in productive reflective knowledge-building.

Across the three conditions in our study, we identified two broad classes of behaviors with distinct learning profiles. Reflective knowledge-building activities involved self-monitoring, knowledge integration, and generation of inferences. As hypothesized, these activities were associated with higher performance on measures of fact recall and comprehension. In contrast, knowledge-telling activities consisted mainly of paraphrasing or summarization of text information, with infrequent elaboration or self-monitoring. These activities showed limited positive association to fact recall, but only weak or negative relationships to deeper understanding.

As in previous work, our peer tutors showed a knowledge-telling bias, generating a high frequency and proportion of knowledge-telling episodes. This is comparable to the results of Fuchs et al. (1997) and King et al. (1998), whose less-trained tutors often generated lecture-like explanations that restated text information. Our study further highlighted a tendency for peer tutors to overtly monitor their own comprehension less than half the time. However, one must acknowledge that the frequency of monitoring and knowledge-building episodes for peer tutors was not negligible. In absolute terms, peer tutors were comparable to self-explainers, and excerpts D and E presented earlier showed instances of peer tutors engaging in elaborated reviewing and sense-making while tutoring a peer. Thus, the peer tutoring experience did indeed contain inherent opportunities for beneficial reflective knowledge-building, although peer tutors did not take advantage of these opportunities as often as one would like.

These findings suggest that educators would benefit by closely studying how knowledge-building opportunities naturally arise during tutoring, and then developing tutoring tactics that specifically address these inherent opportunities. For example, peer tutors often review the material with their pupils to ensure that tutees received and understood important concepts. In our data, we observed that tutors sometimes offered "elaborated reviews" in which they worked to enrich their usual reviewing activities with new insights, connections between concepts, or perspectives. This strategy not only helps tutors extend their knowledge, but likely benefits tutees by exposing them to more and deeper ideas about the topic. Thus, support for tutor learning may not require the sacrifice of tutee learning; both participants can benefit when tutors engage tutees in deep-level discussions about the material (e.g. Chi et al., 2001; Graesser et al., 1995).

Another opportunity occurred when peer tutors encountered ideas they did not fully understand. This may occur regularly for peer tutors whose knowledge is likely to be somewhat incomplete or disorganized. In our study, some tutors spontaneously confronted such knowledge gaps in “sense-making” episodes, in which they used prior knowledge and inferences to revise their understanding. As above, tutors not only deepen their own knowledge through sense-making, they may have help tutees by providing a model of the learning process. Tutees could learn by listening to tutors reason out loud about their errors, and perhaps appropriate similar strategies for themselves.

Commonplace tutor-tutee interactions, particularly tutee questions, also inherently supported reflective knowledge-building. Students who tutored a peer in a face-to-face setting outperformed students who also generated tutorial explanations, but without tutee interaction. This effect was stronger for measures of factual knowledge, but was also noticeable for measures of deeper understanding. These outcomes were paralleled by the finding that peer tutors produced more reflective knowledge-building episodes than tutorial explainers. In both of these conditions, the learners had to generate instructional explanations to convey information to another person. Tutorial explainers received no tutee input and thus had complete control over their explanations. However, peer tutors interacted with a tutee who could influence how the material was discussed. Thus, tutor-tutee interaction seemed to elicit tutors’ knowledge-building.

More specific evidence for this claim came from focused analyses of the tutoring condition. Whereas tutors initiated the majority of knowledge-telling episodes during the tutoring sessions, tutees initiated more knowledge-building episodes. More directly, we found that tutee questions had a meaningful impact on tutors’ explanations. When tutees asked shallow questions about basic facts, tutors gave unelaborated responses exhibiting little self-monitoring. However, deep tutee questions involving inferences offered chances for tutors to further integrate, apply, and generate ideas, and to monitor their own understanding. Thus, deep tutee questions were a specific mechanism by which the tutoring process stimulated peer tutors' reflective knowledge-building. Research has shown that tutees also benefit from asking deep questions (e.g. Graesser & Person, 1994).

What causes the knowledge-telling bias?

Despite the many learning opportunities that can arise, the prevalence of the knowledge-telling bias shows that peer tutors often prefer to deliver rather than develop their knowledge. One goal for future research is to identify the causes of this bias. This will not be easy because many factors might be relevant. For example, younger tutors may not be as developmentally prepared to manage their own learning while tutoring a peer (Wood et al., 1995). Tutors may also need time to practice their skills before they can engage in complex knowledge-building activities (Fuchs et al., 1994). The learning domain could also shape the kinds of knowledge-building that occur. Research in math (e.g. Fuchs et al., 1997) and science domains (King et al., 1998) have yield consistent results, but research that directly compares tutoring and learning across multiple domains, and domains beyond math and science, are needed for conclusive answers. In our concluding sections we briefly consider two potential factors related to our study.

Metacognitive limitations. One cause of the knowledge-telling bias may be tutors' failure to monitor themselves effectively. Self-monitoring is a defining part of reflective knowledge-building, and a breakdown in metacognition could negate many learning opportunities. Research has demonstrated that metacognitive self-monitoring is difficult (Hacker, 1998). Learners often have problems with successful error detection (e.g. Baker, 1979; Glenburg et al., 1982; Maki, 2005; Pressley et al., 1990), failing to notice contradictions between ideas in the text or prior knowledge. Instead, students display an "illusion of knowing" in which they overestimate how well they know the information. Students also have trouble choosing appropriate strategies (e.g. Azevedo & Hadwin, 2005; Hacker, 1998; Palincsar & Brown, 1984). Even when comprehension failures are detected, learners ignore the problem or use shallow strategies that cannot resolve them.

Similar problems likely arise for peer tutors. Tutors may not often monitor their explanations and knowledge if they are overconfident in their domain mastery. When monitoring does occur, tutors may not even realize when their knowledge contradicts the source materials, or when current explanations contradict previous explanations. Moreover, peer tutors may not respond productively when comprehension errors do become apparent. They might simply choose to skip over the difficult concept, thus limiting tutees' exposure to important ideas. Tutors might also simply find the information in a book and read it to the tutee in a knowledge-telling fashion. When tutees become confused by a poor explanation, peer tutors may just reword their current explanation, rather than revising it or generating new elaborations.

Tutors also have the difficult task of monitoring their tutees' understanding (Chi et al., 2004; Foot et al., 1997). It is not clear what barriers this entails for peer tutors' self-monitoring. Tutee monitoring adds to the complexity of the tutors' task, and so tutors may find it challenging to manage both of these processes effectively. This hypothesis may further explain why self-explainers developed a deeper understanding of the material than peer tutors. In both conditions, students had the opportunity to generate explanations, and peer tutors ostensibly had the advantage of being able to participate in interactive dialogue. However, it may be that the quality of metacognitive monitoring differed across the two groups. Peer tutors may have evaluated themselves less efficiently or accurately than self-explainers, leading to less effective reflective knowledge-building.

Our analyses focused on learners' overt metacognitive statements, showing that self-monitoring was linked to higher test scores and knowledge-building. Future research could build upon these analyses by considering specific features of tutors' metacognitive statements. One important variable may be the valence of tutors' judgments, such as positive ("This is easy") or negative ("I don't understand") statements. Another key factor is accuracy, meaning whether tutors' evaluations are predictive of their knowledge of particular concepts. These data would allow us to better assess peer tutors' self-monitoring quality. Are peer tutors indeed overconfident in their domain knowledge? Are their self-evaluations more or less accurate than learners in non-tutoring settings?

In this study, we also examined links between tutee questions and subsequent tutor responses. A similar analysis might be used to assess tutors' reactions to their self-monitoring evaluations. To what extent do knowledge-telling versus knowledge-building activities follow positive or negative judgments? How do tutors' follow-up activities differ when their monitoring is stimulated by a tutee question? A better understanding of how peer tutors respond to their comprehension failures could aid in the design of tutoring techniques that specifically target metacognitive barriers to tutor learning.

Role perceptions. In addition to metacognitive barriers, another source of the knowledge-telling bias may be tutors' role perceptions. Tutoring is not only a cognitive task, but also a role adopted by the tutor (Allen, 1983, Foot et al., 1990). In taking on this role, peer tutors try to enact those behaviors they perceive to be defined by the role. If peer tutors believe that offering emotional support is critical, then they might use tutoring time to discuss the tutees' personal life and stresses. Alternatively, tutors who feel that tutees should stay firmly on-task may avoid all non-academic discussions.

From this perspective, tutors' perceptions could strongly impact tutor learning. One cause of the knowledge-telling bias may be tutors' perception that the goal of tutoring is to convey information in an unmodified form. The source materials contain the key facts, and the tutors' job is only to share these facts with tutees. This may involve putting the ideas into simpler words, but elaboration is not necessary. Generating novel examples or analogies might even be undesirable, because they might confuse the tutee or won't be covered on the exam. A related perception might be that learning from tutoring can (or should) only happen for the tutees. The focus of tutoring is on improving tutee understanding, and so tutors should not waste time pursuing their own curiosity or exploring ideas in too much detail. Indeed, a need to rethink or relearn information while tutoring may be perceived as embarrassing failure to enact the proper tutoring role.

This idea might be best captured in the comparison of self-explainers and tutorial explainers. Participants in both conditions were asked to explain the material and worked alone. However, self-explainers were instructed to explain the material to improve their own understanding, whereas tutorial explainers were told to explain to help another person learn. The differences between the groups were striking. Self-explainers showed better understanding of the material, and engaged in more knowledge-building. One interpretation is that tutorial explainers adopted a "tutor" role focused on transmitting text facts, with little expectation of elaboration or reflection. Peer tutors may have similarly constrained themselves based on role expectations, but tutor-tutee interactions pushed them to occasionally rethink their ideas. In contrast, self-explainers may have adopted a "learner" role in which they were willing to confront their own confusions and explore the information in more depth, without continuous external prompting.

Little is known about students' beliefs about tutoring, but Foot et al., (1990) summarized 8-11 year olds' answers to questions about what they would do if asked to be teachers. The children often reported that they would help others learn facts and skills by explaining and demonstrating ideas, asking them if they understood, and looking for signs of confusion. Foot et al. (1990) also summarized students' ideas about good classroom teachers. Social traits were often mentioned, such as being friendly, patient, or funny. Children also talked about presentation style, such as clarity and enthusiasm. These findings suggest that children perceive both informational (e.g. "give good explanations") and interpersonal (e.g. "be friendly") aspects of the tutoring role. However, more research is needed to determine the relative importance that peer tutors attach to different elements of the role, and how these perceptions influence actual tutoring behaviors. By combining interview or survey data about tutors' role perceptions, and analysis of their actual behaviors, these connections could be explored.

Conclusion

Peer tutoring can be rewarding for both the students who receive tutoring and the tutors. Progress has already been made in developing innovative programs that help tutoring participants achieve meaningful academic gains. The challenge for future research on tutor learning is to further examine the process of how this learning occurs, and perhaps diagnose the barriers to reflective knowledge-building. The results of such work can then be directly applied to the design and redesign of tutoring programs, which will further enhance the strength and scope of the tutor learning effect.

Appendix

Human Visual System Text Excerpts

The entire text is not reproduced here due to the length, but two excerpts are provide to offer a flavor of the writing style and level. These excerpts are chosen to align with two of the episode examples shared, so the reader can further appreciate how text information was used differently in knowledge-telling and knowledge-building episodes. Text Excerpt 1 (from page 7 of our text) describes the lens, ciliary muscles, and the process of accommodation. Episode Excerpt C shows an example of knowledge-building based on this information. Text Excerpt 2 (from page 16 of our text) describes the horizontal and amacrine cells of the retina. Episode Excerpt A shows an example of knowledge-telling based on this text.

(1) Although most of the bending of the light, called refraction, is done by the cornea, some additional focusing is accomplished by varying the thickness of the lens. This process is called accommodation. In this process, the ciliary muscles tighten and thereby increase the refractive power, and thus the focusing power, of the eye. Accommodation enables us to keep an image on the retina sharp as we look at objects located at different distances.

(2) There are two other types of retinal cells called horizontal cells and amacrine cells. Horizontal cells connect receptors to other receptors. Amacrine cells connect ganglion cells to other ganglion cells and bipolar cells to other bipolar cells. These cells do not transmit signals toward the brain, but instead transmit and pool signals laterally across the retina. Thus, neural signals flow both directly toward the brain and laterally across the retina before going to the brain. (from p. 16 of the text materials)

Definitions Test Items

Thirteen components were targeted on the eyeball diagram: aqueous humor, blind spot, choroid, cornea, eyelid, fovea, iris, lens, optic nerve, pupil, retina, sclera, and vitreous humor. An additional 13 terms were included as filler terms: amygdala, capillary, cochlea, duodenum, femur, larynx, ligament, macula, microvillae, nephron, pyramidal cell, sciatic nerve, and valve. On the retina diagram, seven components were targeted: amacrine cell, bipolar cell, cone receptor, ganglion cell, horizontal cell, optic nerve, and rod receptor. An additional nine items were included as filler terms: aqueous humor, barorecptor, blind spot, duodenum, leukocyte, microvillae, nephron, optic chiasm, and pyramidal cell. On both diagrams, some filler terms were unrelated to the eye (e.g. femur) and others were related to the eye but not indicated on the diagram (e.g. macula).

Questions Test Items

The following questions were used in the Questions Test. Questions are listed in the order they appeared; those labeled with an asterisk appeared only on the post-test.

1) Which receptors are responsible for color and detail vision in high levels of light?

2) How is light entering the eye adjusted and directed toward the retina?

3) What is meant by the term “pooling of receptors”?*

4) What are the most direct and least direct pathways that neural signals can take the reach the brain from the receptors?

5) What muscles are involved in the process of accommodation?*

6) Why is the fovea the most sensitive portion of the retina for detecting light patterns?

7) What is the iris? What function does it serve?

8) Color blindness is a disorder in which a person is unable to see colors. Some people with color blindness are “blind” to all colors, while others are “blind” to just one or two colors. What might be the problem underlying color blindness? How could a person be “blind” to just one or two colors?

9) What is refraction?

10) What are some possible reasons why a person may need glasses? How do glasses help to compensate for these problems?

11) Why do humans see better in the light than the dark?

12) What structure us responsible for over two-thirds of the eye’s focusing power?

13) Most people never notice the “blind spot” in their vision? Why don’t people notice?*

14) What different types of cells are bipolar cells connected to?

15) Some animals can see in the dark. How might their eyes be structured different than human eyes so that this is possible?

16) What is the blind spot? Where is it located?*

17) How are receptors, neurons, and blood vessels arranged in the retina?*

18) What are the two main purposes of the vitreous humor?*

19) Diabetes is an illness in which the body does not produce enough insulin (a hormone needed to metabolize sugar). If untreated, diabetes can sometimes lead to swollen blood vessel and blindness (along with other health problems. How could diabetes lead to blindness?*

20) What happens when a molecule of photopigment absorbs a photon?*

References

Allen, V.L. & Feldman, R.S. (1976). Studies on the role of tutor. In V. Allen, ed, *Children as teachers: Theory and research on tutoring*, pp. 113-129. New York: Academic Press.

Allen, V.L. (1983). Impact of the role of the tutor on behavior and self-perceptions. In J. Levine & M. Wang (Eds.), *Teacher and student perception: Implications for learning* (pp. 367-389). Hillsdale, NJ: Elrbaum.

Annis, L. (1983). The processes and effects of peer tutoring. *Human Learning* 2: 39-47.

Atkinson, R., Renkl, A., Merrill, M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774-783.

Azevedo, R. & Hadwin, A. (2005). Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science, 33*, 367-379.

Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior, 11*(4), 367-374.

Bargh, J.A. & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology* 72: 593-604.

Benware, C.A. & Deci, E.L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research* Journal 21(4): 755-765.

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.

Brown, A. & Day, J. (1983). Macrorules for summarizing text: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*, 1-14.

Chan, C. (2001). Peer collaboration and discourse patterns in learning from incompatible information. *Instructional Science, 29,* 443-479.

Chi, M. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences* 6(3): 271-315.

Chi, M. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser, ed , *Advances in instructional psychology*, pp. 161-238. Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Chi, M., deLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanation improves understanding. *Cognitive Science* 18: 439-477.

Chi, M., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R.G. (2001). Learning from human tutoring, *Cognitive Science* 25: 471-533.

Cohen, J. (1986). Theoretical considerations of peer tutoring. *Psychology in the Schools* 23: 175-186.

Cohen, P.A., Kulik, J.A., & Kulik, C.C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal* 19(2): 237-248.

Coleman, E.B. (1998). Using explanatory knowledge during collaborative problem-solving in science. *Journal of the Learning Sciences* 7: 387-427.

Coleman, E.B., Brown, A.L., & Rivkin, I.D. (1997). The effect of instructional explanations on formal learning from scientific texts. *The Journal of the Learning Sciences* 6(4), 347-365.

Davis, E. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences*, *12*(1), 91-142.

Derry, S.J. & Potts, M.K. (1998). How tutors model students: A study of personal constructs in adaptive tutoring. *American Educational Research Journal* 35(1): 65-99.

Duit, R., Roth, W., Komorek, M., & Wilbers, J. (2001). Fostering conceptual change by analogies – between Scylla and Charybdis. *Learning and Instruction, 11*, 283-303.

Fantuzzo, J.W., Riggio, R.E., Connelly, S., & Dimeff, L.A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A componential analysis. *Journal of Educational Psychology* 81(2): 173-177.

Flavell, J. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34,* 906-911.

Foot, H., Shute, R., Morgan, M., & Barron, A. (1990). Theoretical issues in peer tutoring. In H. Foot, M. Morgan, & R. Shute (Eds.), *Children Helping Children.* (pp. 65-92). New York: Wiley.

Fogarty, J. & Wang, M. (1982). An investigation of the cross-age peer tutoring process: Some implications for instructional design and motivation. *Elementary School Journal*, *82*(5), 450-469.

Fox, S. (1996). *Human Physiology, 5th Ed.* Chicago: WCB Publishers.

Foot, H., Shute, R., Morgan, M. (1997). Children's sensitivity to lack of understanding. *Educational Studies, 23*(2), 185-194.

Fuchs, L.S., Fuchs, D., Bentz, J., Phillips, N.B., & Hamlett, C.L. (1994). The nature of students' interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal* 31: 75-103.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal* 97(3), 223-249.

Gartner, A., Kohler, M.C., & Riessman, F. (1971). *Children teach children: Learning by teaching*. New York: Harper & Row.

Glenburg, A., Wilkinson, A., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory and Cognition*, 10(6), 597-602.

Goldstein, E.B. (1999). *Sensation and perception*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Graesser, A.C. & McMahen, C. (1993). Anomalous information triggers questions when adults solve quantitative problems and comprehend stories. *Journal of Educational Psychology, 85*(1), 136-151.

Graesser, A.C. & Person, N.K. (1994). Question asking during tutoring. *American Educational Research Journal* 31(1): 104-137.

Graesser, A.C., Person, N.K., & Magliano, J.P. (1995). Collaborative Dialogue Patterns in Naturalistic One-to-One Tutoring. *Applied Cognitive Psychology* 9: 495-522.

Greenwood, C.R., Carta, J.J., & Hall, R.V. (1988). The use of peer tutoring strategies in classroom management and educational instruction. *School Psychology Review* 17(2): 258-275.

Hacker, D. (1998). Self-regulated comprehension during normal reading. In. D. Hacker, J. Dunlosky, & A. Graesser, (Eds.), *Metacognition in educational theory and practice* (pp. 165-191. Mahwah, NJ: Erlbaum.

Hamaker, C. (1986). The effects of adjunct question on prose learning. *Review of Educational Research*, 56(2), 212-242.

Ismail, H. & Alexander, J. (2005). Learning within scripted and non-scripted peer-tutoring session: The Malaysian context. *Journal of Educational Research*, 99(2), 67-77.

Juel, C. (1996). Learning to learn from effective tutors. In L. Schauble & R. Glaser, eds, *Innovations in learning: New environments for education*, pp. 49-74. Mahwah, NJ: Lawrence Erlbaum Associates.

King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal* 31(2): 338-368.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review, 10*(1), 57-74.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology* 90(1): 134-152.

Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook of research on teaching: 4th Ed.* (pp. 333-357). Washington, DC: AERA.

Maki, R., Shields, M., Wheeler, A., & Zacchilli, T. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*(4), 723-731.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction* 7(3): 197-244.

McNamara, D., Louwerse, M., Cai, Z., & Graesser, A. (2005, January 1). Coh-Metrix version 1.4. Retrieved September, 19, 2006, from http://cohmetrix.memphis.edu.

Merrill, D., Reiser, B., Merrill, S., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction* 13(3): 315-372.

Osborne, R. & Wittrock, M. (1983). Learning science: A generative process. *Science Education, 67*(4), 489-508.

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

Peverly, S. & Wood, R. (2001). The effects of adjunct questions and feedback on improving the reading comprehension skills of learning disabled adolescents. *Contemporary Educational Psychology*, 26, 25-43.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research* 31: 459-470.

Pressley, M., Ghatala, E. S., Woloshyn, V., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly, 25*(3), 232-249.

Pressley, M., McDaniel, M., Turnure, J., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(2), 291-300.

Putnam, R. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal* 24(1): 13-48.

Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction* 5: 21-36.

Rohrbeck, C.A., Ginsburg-Block, M.D., Fantuzzo, J.W., & Miller, T.R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Eductional Psychology* 95(2): 240-257.

Roy, M. (2001). *Comprehension and learning through multimedia: Integrative processing of text and illustrations.* Doctoral dissertation, McGill University, Montreal, Quebec, Canada.

Saulny, S. (2005, April 4). A lucrative brand of tutoring grows unchecked. *The New York Times,* p. A18.

Scruggs, T.E. & Osguthorpe, R.T. (1986). Tutoring interventions with special education settings: A comparison of cross-age and peer tutoring. *Psychology in the Schools* 23: 187-193.

Sharpley, A.M., Irvine, J., & Sharpley, C.F. (1983). An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal* 20(1): 103-111.

van der Meij, H. (1994). Student questioning: A componential analysis. *Learning and Individual Differences, 6*(2), 137-161.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning doing human tutoring? *Cognition and Instruction,* 21(3): 209-249.

Webb, N. (1989). Peer instruction, problem-solving, and cognition: Multidisciplinary perspectives. *International Journal of Educational Research*, *13*, 21-39.

Wood, D., Wood, H., Ainsworth, S., & O'Malley, (1995). On becoming a tutor: Toward and ontogenetic model. *Cognition and Instruction*, *13*(4), 565-58.

Table 1

Distribution of demographic factors by condition

Demographic Factor	Peer Tutors	Peer Tutees	Tutorial Explainers	Self-Explainers	Overall Mean/Total
	Conditions				
Mean Age	24.9 (7.3)[a]	19.9 (1.4)	19.0 (1.2)	19.0 (1.1)	20.7 (4.4)
Mean VSAT	610 (84.3)	620 (82.3)	560 (56.8)	640 (56.8)	608 (74.7)
Mean GPA	3.2 (.61)	3.5 (.35)	3.3 (.37)	3.4 (.34)	3.3 (.44)
Gender					
Female	6	6	5	5	22
Male	4	4	5	5	18
Year					
Freshman	0	1	4	4	9
Sophomore	3	2	4	3	12
Junior	2	3	1	3	9
Senior	5	4	1	0	10
Major					
Science	5	6	6	4	21
Non-sci.	5	4	4	6	19

Note. Age, VSAT scores, and GPA are reported as means with standard deviations in parentheses. Gender, year, and major are reported as frequency counts.

[a]All differences are non-significant (p>.05) except for age. The mean age of Peer Tutors was higher than other conditions due to one much older student, $F(3,36) = 5.5$, $p < .01$.

Table 2

Mean Scores on the Definitions Test and Questions Test

	Condition			
Assessment	Peer Tutors (T)	Tutorial Explainers (E)	Self Explainers (S)	Significant Pair-wise Comparisons
Definitions Test				
Pre-test	6.4 (6.1)	4.1 (2.7)	6.1 (3.3)	
Post-reading	24.7 (14.1)	18.6 (6.7)	24.5 (10.1)	
Post-test	33.3 (10.2)	21.3 (4.8)	37.6 (11.4)	T > E, $ES = 1.51^a$; S > E, $ES = 1.86^b$
Questions Test				
Post-reading	19.8 (6.1)	13.2 (5.1)	19.8 (5.3)	T > E, $ES = 1.17^a$; S > E, $ES = 1.27^a$
Post-test	26.9 (4.9)	24.1 (5.3)	32.4 (4.9)	S > T, $ES = 1.13^a$; S > E, $ES = 1.64^b$
Repeat	20.5 (3.1)	17.8 (3.3)	23.1 (3.1)	
New	6.4 (2.8)	6.3 (3.0)	9.3 (2.8)	

Note. Values in parentheses indicate standard deviations. Scores reported for Post-test scores are adjusted means. $^a p < .05$, $^b p < .01$.

Table 3

Correlations Between Mean Frequency of Episodes and Post-Test Scores

	Definitions Test	Questions Test
Knowledge-Telling Episodes	.25	-.12
Concept Elaboration	.38[a]	.15
Knowledge-Building Episodes	.58[b]	.47[b]
Elaborated Reviewing	.50[b]	.49[b]
Sense-making	.37[a]	.38[a]
Metacognitive Activity		
With Monitoring	.59[b]	.42[a]
Without Monitoring	.13	-.22

Note. [a]$p < .05$. [b]$p < .01$.

Table 4

Patterns of Episodes Across Conditions

	Condition			
	Peer	Tutorial	Self	
Episode Category	Tutors (T)	Explainers (E)	Explainers (S)	Significant Pair-wise Comparisons
	Frequency Data			
Total Episodes	40.2 (9.5)	25.5 (12.5)	34.0 (9.5)	T > E, *ES* = 1.32[a]
Knowledge-Telling	29.1 (4.9)	21.8 (10.8)	20.4 (8.1)	
Concept Elaboration	6.8 (5.1)	3.3 (2.2)	4.9 (3.7)	
Knowledge-Building	11.1 (6.6)	3.7 (3.3)	13.6 (5.1)	T > E, *ES* = 1.42[a]; S > E, *ES* = 2.30[b]
Elaborated Reviewing	1.5 (1.4)	.20 (.42)	2.4 (1.4)	S > E, *ES* = 2.09[c]
Sense-making	3.8 (2.0)	.40 (.84)	3.1 (3.5)	T > E, *ES* = 2.22[a]; S > E, *ES* = 1.06[c]
Metacognitive Activity				
With Monitoring	21.1 (5.1)	6.7 (4.6)	16.9 (9.4)	T > E, *ES* = 2.96[b]; S > E, *ES* = 1.38[b]
Without Monitoring	27.8 (11.7)	20.1 (11.0)	18.2 (4.5)	
	Proportion Data			

Knowledge-Telling	.74 (.10)	.87 (.11)	.60 (.16)	E > S, *ES* = 1.98[b]; T > S, *ES* = 1.04[c]
Without Monitoring	.55 (.12)	.75 (.18)	.55 (.17)	

Note. Values in parentheses indicate standard deviations. [a]$p < .05$, [b]$p < .01$, [c]$p < .06$