OPPORTUNITES AND BARRIERS FOR TUTOR LEARNING: KNOWLEDGE-BUILDING, METACOGNITION, AND MOTIVATION

by

Rod David Roscoe

Bachelor of Science in Psychology and Neuroscience, University of Pittsburgh, 1999

Master of Science in Psychology, University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

the University of Pittsburgh in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Rod David Roscoe

It was defended on

August 16th, 2007

and approved by

Janet Schofield, Ph.D, Professor of Psychology

Christian Schunn, Ph.D., Associate Professor of Psychology

Kurt VanLehn, Ph.D., Professor of Computer Science

Dissertation Advisor: Michelene Chi, Ph.D, Professor of Psychology

OPPORTUNITES AND BARRIERS FOR TUTOR LEARNING: KNOWLEDGE-BUILDING, METACOGNITION, AND MOTIVATION

Rod David Roscoe, Ph.D.

University of Pittsburgh, 2007

Peer tutoring is an educational intervention in which students tutor other students. An important finding from peer tutoring research is that tutors can learn by tutoring. This *tutor learning effect* applies across tutoring formats, student populations, and domains. Unfortunately, the average magnitude of these gains is underwhelming. This finding may arise because peer tutors do not often engage in *knowledge-building* activities as they teach (i.e. self-monitoring, integrating new and prior knowledge, and generating new ideas) which are associated with stronger tutor learning outcomes. Instead, peer tutors display a *knowledge-telling bias* by primarily summarizing the materials with little elaboration or self-monitoring.

A critical goal for tutor learning research is to understand the sources of this bias. A metacognitive hypothesis is that tutors do not adequately monitor their understanding, thus preventing them from recognizing and revising comprehension failures. A motivational hypothesis is that tutors choose less productive strategies because they possess negative attitudes towards the material or tutoring task. These hypotheses were assessed using objective assessments of tutor learning, coding of tutors' behaviors at multiple grain-sizes, and self-report measures of self-efficacy and interest. Previous findings were replicated to show that reflective knowledge-building activities were associated with significantly higher post-test scores. Peer tutors also showed a clear knowledge-telling bias by primarily generating unelaborated summaries and reviews of the material, which were not associated with higher scores.

Mixed support was found for the metacognitive hypothesis. Although self-monitoring was positively associated with knowledge-building, high and low-performing tutors did not differ in their overall self-monitoring, nor in specific kinds of self-monitoring statements. However, high-performing tutors' self-monitoring was more likely to occur in conjunction with elaboration of the material. Clearer support was found for the motivational hypothesis. Tutors' interest and self-efficacy were positively associated with test scores and more frequent reflective knowledge-building. Thus, peer tutors' decisions about to teach and think about the material were seemed to be influenced by their attitudes. Suggestions for designing tutoring programs to support interest and self-efficacy are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

There are many people and institutions I would like to thank for their encouragement and support. The university itself has been extremely generous to me in the form of scholarships and fellowships, some of which directly funded my dissertation research. I also greatly appreciate the fine community of scholars and friends I have found in Department of Psychology and Learning Research and Development Center. In working with the LRDC and departmental diversity committees, I have been impressed by the level of commitment our community has shown to support and enrich all students. Special thanks are also in order for my dissertation committee members. I have learned so much from them not only in regards to this project, but also many other academic endeavors: Janet Schofield, Chris Schunn, and Kurt VanLehn.

Working in the lab of Dr. Michelene Chi for several years, I have had the opportunity to learn from a number of great colleagues: Kirsten Butcher, Scotty Craig, Randi Engle, Soniya Gadgil, Bob Hausmann, Mark McGregor, Bibinaz Pirayesh, Marguerite Roy, Stephanie Siler, and Roger Taylor. Throughout all the ups and downs of academic life, they have been generous with their expertise and words of encouragement. The center of the lab is Micki herself, who has provided measureless guidance in her role as my academic advisor. Although her time is in great demand by scholars around the world, she is almost always available to share her valuable

expertise and insights. I have benefited from countless impromptu visits to her office with "quick questions" that turned into much longer discussions about research and academia.

Heartfelt thanks also go out to my many friends who have been an integral part of "surviving" graduate school, and helping me to keep a balanced perspective and a sense of humor. This list includes all of my lab members, past and present, and many others both in and out of school. There are simply too many names to include in this small space… but anyone who has made me laugh during my time in graduate should know how important you are to me.

Finally, words cannot describe my gratitude and love for my family: Jane Roscoe, Rodney Roscoe, and Craig Stewart. They have been my most tireless advisors, champions, and cheering squad. No matter what I have needed, I knew that I could always count on them. I can only hope to live up to their expectations of me and be worthy of their generosity and wisdom. I simply could not have done it without them!

1.0 INTRODUCTION

Human tutoring is one of the most effective educational interventions, with tutored students often performing significantly better than students in classroom settings (Bloom, 1984). Although it would be ideal if every student could receive expert tutoring, this turns out to be implausible due to the expense and limited availability of expert tutors. One popular alternative is peer tutoring, in which students are recruited as instructors of other students (Greenwood, Carta, & Hall, 1988). Student tutors are plentiful, and may be recruited as volunteers, compensated with course credit, or hired for a modest wage. More importantly, evidence shows that student tutors can be successful instructors, with an average Cohen's d effect size (*ES*) of about .40 (Cohen, Kulik, & Kulik, 1982). Although these gains do not surpass those of expert tutoring (Bloom, 1984), peer tutoring offers an additional academic benefit. Student tutors can develop a better understanding of the subject matter as a result of their teaching experiences (Cohen et al., 1982; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003), referred to as the *tutor learning effect*.

Tutor learning benefits seem to apply quite broadly. For example, one basic feature of tutoring programs is the format. In cross-age tutoring, older and more advanced students instruct younger novices. These roles typically remained "fixed" because tutees cannot teach the more advanced tutors. In contrast, same-age tutoring involves tutors and tutees who are of a similar age and grade. The nearness in age and background means that tutors and tutees may have overlapping or complimentary knowledge. As a result, same-age tutoring roles can be fixed, but are often “reciprocal.” The students take turns tutoring each other, and benefit from both tutoring and being tutored. Despite such variations, tutor learning effects have been demonstrated in all formats, such as cross-age tutoring with a small age-gap (Sharpley, Irving, & Sharpley, 1983) or a large age gap (Juel, 1996), same-age tutoring with fixed roles (Coleman, Brown, & Rivkin, 1997), and same-age tutoring with reciprocal roles (Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997; King, Staffieri, & Adelgais, 1998).

Another feature of tutoring programs is the demographics of the participating students. Tutor learning has been observed in all age groups, including college (Annis, 1983), high school, (Cloward, 1967), middle school (King et al., 1998), and elementary school students (Sharpley et al., 1983). Some researchers have studied tutor learning in special populations that often experience academic challenges. These studies have shown positive outcomes from tutors from underprivileged backgrounds (Rohrbeck et al., 2003), minority students (Robinson, Schofield, & Steers-Wentzel, 2005), and students with learning disabilities or behavior disorders (Cook, Scruggs, Mastropieri, & Casto, 1986; Mathes & Fuchs, 1994).

The finding that students can learn by tutoring has provided a persuasive justification for the implementation of peer tutoring programs (Allen & Feldman, 1976; Gartner, Kohler, & Riesmann, 1971). However, a closer look at the literature raises some questions about the validity of this assumption. Although the *scope* of the tutor learning effect is impressive, the *magnitude* of the effect is more problematic. In a recent review of tutor learning research, Roscoe & Chi (in press-B) summarized the effect sizes (Cohen’s d) reported by a number of prior meta-analyses. Most effects clustered around a value of .35, which indicated a positive but low-to-moderate effect (Cohen, 1992). Although there is ample evidence that tutors can learn, such outcomes are not guaranteed and tutor learning gains may not be very strong.

Roscoe & Chi (in-press-B) attributed this finding to a pervasive pattern of tutor instructional activities that they termed the "knowledge-telling bias." Peer tutors relied heavily on tutoring by summarizing the source materials and giving didactic explanations with little elaboration, deeper reasoning, or metacognitive self-monitoring. These knowledge-telling behaviors were linked to lower scores on learning assessments. Peer tutors learned best when their tutoring behaviors incorporated "reflective knowledge-building," including monitoring one's own comprehension, integrating new and prior knowledge, and generating new ideas through inferences and reasoning (Fuchs et al. 1997; Ismail & Alexander, 2005; King et al.1998; Roscoe & Chi, in-press-A). However, these productive behaviors occurred rarely unless extensive tutor training was provided.

Although there is little doubt in the literature about *what* peer tutors need to do to learn, much less is known about *why* peer tutors choose to engage in particular tutoring activities. The knowledge-telling bias is likely to be influenced by a number of issues, such as task perceptions, self-perceptions, prior knowledge, metacognitive skills, or tutor-tutee interactions. A critical goal for research on peer tutoring is to identify the factors or environments that contribute peer tutors' knowledge-telling bias (Roscoe & Chi, in-press-B). Such research could be used to improve tutor learning by designing training methods that specifically target commonplace obstacles to tutor learning. Thus, a better understanding of the knowledge-telling bias could lead to dramatic gains in tutor learning across many different programs.

The current project further explores the role of knowledge-building in tutor learning, and considers several potential sources the knowledge-telling bias. First, I review prior research on the role knowledge-building and knowledge-telling in learning from explaining and responding to tutee questions. Two plausible sources of the knowledge-telling bias are then proposed. One hypothesis is that peer tutors do not engage in knowledge-building because they do not monitor their own understanding adequately. Another hypothesis is that tutors avoid knowledge-building because they lack self-efficacy or interest for discussing difficult topics or thinking deeply. These metacognitive and motivational hypotheses are seen as complementary, not competing.

These hypotheses are then evaluated in a laboratory study on tutor learning in which college students studied a text about the human visual system and then taught the material to another student. Tutor knowledge and learning were assessed at several time points, and tutoring sessions were videotaped and transcribed to analyze tutors' learning behaviors. Learning assessment data and verbal data were also collected for a group of students who read and reviewed the text on their own using a think-aloud strategy. These data enable comparisons of learning and behavior between tutors and non-tutors. I conclude by discussing the implications of the results and directions for future research.

1.1 PROCESSES OF TUTOR LEARNING

Analyses of expert or experienced tutors have demonstrated the complex nature of the tutoring task. Effective tutoring can involve a variety of tactics, such as explaining, questioning, assessment, and feedback (Lepper, Drake, & O'Donnell-Johnson, 1997; McArthur, Stasz, & Zmuidzinas, 1990; Merrill, Reiser, Merrill, & Landes, 1995; Shah, Evens, Michael, & Rovick, 2003; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003).

Educators have hypothesized that tutor learning is a direct result of tutors' engagement in these common tutoring behaviors (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, these intermingled processes of knowledge construction and metacognition are referred to as reflective knowledge-building (Roscoe & Chi, in-press-A, in-press-B).

This hypothesis can be made more concrete by considering the opportunities for reflective knowledge-building in two ubiquitous tutoring activities, explaining the material to the tutee and responding to tutee questions.

1.1.1 Explaining and reflective knowledge-building

Tutors frequently use verbal explanations to communicate key ideas, principles, and relationships, and to correct tutee errors. Such explanations may involve developing analogies, summarizing main ideas, working through examples, etc., and can be used to share known information or make sense of new information. Not surprisingly, the richness of explaining has led many to hypothesize that it contributes to tutor learning (Bargh & Schul, 1980; Cohen, 1986; Coleman et al., 1997; Gartner et al., 1971; King, 1998). Simply generating and articulating these explanations may improve recall of the material (Osborne and Wittrock 1983), and similar benefits may arise from summarizing main ideas (e.g. Brown and Day 1983). Research has also shown that students can gain a deeper understanding by generating analogies (Duit, Roth, Komorek, & Wilbers, 2001), and by decomposing and explaining worked-out examples (Atkinson, Renkl, & Merrill, 2003; Chi, Bassok, Lewis, Reimann, & Glaser, 1989).

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). 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 meta-memory, which involve evaluation of one's own knowledge and understanding (e.g. Flavell 1979; Hacker 1998; Maki, Shields, Wheeler, & Zacchilli, 2005). As students progress through a learning task, they can judge themselves by various criteria (e.g. feelings of confusion, ability to recall facts, 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. 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 explanations may require tutors to reorganize their own flawed mental models by rearranging conceptual connections. Similarly, striving to create explanations that 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 (Chi, 2000). These processes should result in a more organized, correct, and/or deeper understanding of the material (Bargh and Schul 1980; Chan, Burtis, & Bereiter, 1997; Coleman et al. 1997; King et al. 1998; Roscoe & Chi, in-press-A).

In many ways, reflective knowledge-building may be analogous to self-explaining, in which learners make sense of new information by using prior knowledge and inferences to explain it to themselves (Chi, 2000). These self-explanations occur when learners realize they do not understand or cannot solve a problem, perhaps because they are missing information or a have a flawed mental model (Chi, 2000; Chi & VanLehn, 1991). Such "impasses" may occur frequently for peer tutors who are trying to explain ideas they have learned only recently or imperfectly. In these cases, tutors may have to work out a correct explanation for themselves before (or while) conveying this knowledge to the tutee.

Research on self-explaining has established the efficacy of this strategy for learning in a variety of contexts, such as reading a text (Chi, deLeeuw, Chiu, & LaVancher, 1994; McNamara, 2004), solving problems (Chi et al., 1989; Renkl, 1997), or learning with multiple representations (Ainsworth, 1999). Peer tutors' explanations may often be embedded in such tasks. For example, chemistry tutors might walk tutees through examples of reactions to teach how reagents interact and energy is released. Tutors may use formulas to calculate energy, the periodic table, and drawings of molecular bonds to represent the reaction processes. To fully utilize these tools, peer tutors may need to deconstruct example principles and steps, and explain how information across representations is utilized, perhaps encouraging a significant amount of self-explanation.

Only a small number of studies have examined peer tutors' explanations, and fewer studies have analyzed such behaviors in conjunction with tutor learning data (Roscoe & Chi, in-press-B). However, the results of such work are fairly consistent. Peer tutors learn most effectively when their explanations incorporate reflective knowledge-building, but peer tutors show a strong bias towards relying on knowledge-telling explanations.

Several studies have described tutors' behaviors in naturalistic (i.e. no training) settings with a focus on tutee learning (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Fogarty and Wang, 1982; Graesser and Person, 1994; Graesser, Person, & Magliano, 1995. These studies did not collect tutor learning data, but provided insights about tutors' typical behaviors. A common finding was that explaining was a central activity for tutors of all ages. However, the 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.

Other studies have examined tutors' explanations in conjunction with tutor learning. 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). These results suggest that tutors benefited from trying to give explanations that contained new insights and integrated the text concepts through inferences. However, tutorial dialogues were not directly analyzed, and so it was not clear whether tutors in either group actually engaged in reflective knowledge-building or not.

Fuchs et al. (1997) examined math tutors' explaining behaviors more directly. They found that elementary school reciprocal peer tutors trained to give ''conceptual explanations'' in math (e.g. explain underlying concepts and create examples) outperformed tutors who were only told to give feedback when tutees made mistakes. The conceptual explanation tutors also produced more knowledge-integration statements and asked more problem-solving questions. The feedback-only 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 generated explanations that incorporated knowledge-building elaborations seem to learn more effectively than tutors who only summarized (Coleman et al. 1997; Fuchs et al. 1997). However, it was not clear when or why such elaborations occurred, and these studies did not assess how learning from explaining involved tutors' self-monitoring. Another issue was that untrained peer tutors seemed to adopt a knowledge-telling bias. Instead of building upon their knowledge, tutors often ‘‘told what they knew’’ by summarizing and giving didactic lectures (e.g. Chi et al. 2001; Graesser et al. 1995; Fuchs et al. 1997), although some untrained tutors spontaneously engaged in productive activities (Graesser et al. 1995).

1.1.2 Responding to tutee questions and reflective knowledge-building

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

From a tutor learning perspective, tutees' questions may be beneficial in several ways. In one case, tutors' knowledge and explanations may be 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 reviewers' knowledge and a text (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 and Schul 1980; Cohen 1986; Foot, Shute, Morgan, & Barron, 1990; King et al. 1998). Tutee questions may actually stimulate a meaningful proportion of tutors' learning through elaboration and self-monitoring.

The nature and quality of tutees' questions may be an important factor. Shallow questions about basic facts may not elicit deeper responses. Peer tutors most likely have a decent grasp of the material at that level, and so these questions offer little challenge or need for knowledge-building. In contrast, questions that tap deeper reasoning and application may provide more knowledge-building opportunities. These questions are probably more likely to touch upon disorganized, fragmented, or misconceived areas of the tutors' subject matter knowledge (King, 1998; Roscoe & Chi, in-press-A)

As with explanations, few studies have examined the role of questions in tutor learning. Graesser and 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. This result is notable because the knowledge-telling bias persisted even when tutors were trained to give deeper explanations. In contrast, the questioning groups asked more and deeper questions, and also produced more explanation statements that integrated and applied the material, and that prompted critical thinking.

These results were recently replicated by Ismail and Alexander (2005), using high school students enrolled in a physics course. However, neither of these studies 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. It is also important to remember that these studies involved reciprocal tutoring. Tutors both gave explanations in response to questions and received explanations in response to their own questions. Thus, the benefits of each activity are conflated.

Only one study has directly examined how tutee questions influenced peer tutors' knowledge-building, and the role of metacognitive self-monitoring in tutor learning (Roscoe & Chi, in-press-A). Participants were undergraduates with low prior domain knowledge who received no tutor training. The participants first learned about the human eye by studying a text, and then either taught this material to another student (peer tutoring) or generated a videotaped explanation that could be later used by a peer to learn (tutorial explaining). Thus, in the peer tutoring condition tutees could influence tutors' explanations via questions and comments. Tutor and tutee roles were not reciprocal. Post-tests showed that peer tutors were better able to define key terms and answer comprehension questions than were tutorial explainers. Thus, interactions with a tutee seemed to support stronger learning.

Analyses of peer tutors' explanations showed that all participants mainly generated knowledge-telling explanations in which they paraphrased and summarized text facts. These explanations were not significantly correlated with post-test scores. However, peer tutors engaged in significantly more reflective knowledge-building. When reviewing concepts, some peer tutors built upon prior knowledge by generating new inferences and elaborations. Tutors also sometimes made sense of confusing concepts by reasoning aloud in a manner akin to self-explaining. These elaborated reviewing and sense-making activities were highly metacognitive (i.e. contained overt self-monitoring statements), and were significantly and positively correlated with fact recall and comprehension measures. Thus, tutor-tutee interactions seemed to support learning by stimulating reflective-knowledge-building. Indeed, subsequent analyses further showed that many of tutors' knowledge-building activities were initiated by tutee questions and comments. Knowledge-telling activities were primarily tutor-initiated.

The impact of tutee questions were directly assessed by identifying question-response exchanges (i.e. tutee questions and tutor responses). Tutee questions were coded as "deep" or "shallow" depending on whether they contained or required an inference. Tutor responses were also coded as deep or shallow, and as "metacognitive" or "non-metacognitive" based on whether they contained overt self-monitoring statements (e.g. "I'm not sure what that means"). Shallow questions were much more common than deep questions, but both occurred. Results showed that deep questions were significantly more likely to elicit deep and metacognitive responses, whereas shallow questions elicited shallow and non-metacognitive answers.

This pattern was not perfect. Tutors offered shallow replies to deep questions about 30% of the time. Perhaps tutors felt the answer was obvious, and only a short reply (e.g. "That makes sense") was needed. It is also possible that tutors did not know the answer and chose not to try to answer it (e.g. "I dunno. The book doesn't say"). Thus, the knowledge-telling bias might also influence how tutors respond to questions. Due to a relatively small sample size, relationships between different question-response patterns and learning could not be assessed.

In sum, available research offers 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 and learn more effectively (Chi et al. 2001; King et al. 1998; Ismail and Alexander 2005; Roscoe & Chi, in-press-A). These studies also provide further evidence of the knowledge-telling bias and its robustness. Untrained tutors in these studies preferred to give knowledge-telling explanations, and even training to give deeper explanations was not always sufficient to overcome the bias (King et al. 1998). Although deep tutee questions can help to overcome this bias, some tutors found ways to ignore or evade these opportunities (Roscoe & Chi, in-press-A).

1.2 POTENTIAL SOURCES OF THE KNOWLEDGE-TELLING BIAS

There is little question in the peer tutoring literature that when peer tutors' instructional activities incorporate reflective knowledge-building, learning outcomes are stronger. However, despite the rich learning opportunities inherent to the tutoring process, tutors prefer to deliver rather than develop their knowledge. This widespread knowledge-telling bias creates barriers that prevent peer tutors from achieving the kinds of strong learning gains that are known to be possible.

Extensive training in particular strategies is one way to overcome the knowledge-telling bias. Peer tutors can be given detailed instructions on how to tutor, weeks of practice using the strategies, and perhaps rewards for using the strategies correctly. This approach has certainly yielded a variety of programs that produce measurable learning gains (King, 1998; McMaster, Fuchs, & Fuchs, 2006; Rohrbeck et al. 2003; Topping & Ehly, 2001). However, this method may ignore underlying causes of the knowledge-telling bias. If one assumes that tutors use poor strategies only because they do not know any better, then this training approach makes sense. However, it is more likely that peer tutors' choices to engage in knowledge-telling versus knowledge-building are determined by more than just their familiarity with alternative strategies. Indeed, some studies have shown that even untrained tutors can spontaneously engage in reflective knowledge-building (Roscoe & Chi, in-press-A), and trained tutors do not always completely or effectively adopt the trained strategies (Dufrene, Noell, Gilbertson, Duhan, 2005).

Research on the sources of the knowledge-telling bias can help us to identify the barriers to reflective knowledge-building and tutor learning, and design training methods that target these problems directly. Future training programs may benefit from adopting a "two-pronged approach" that simultaneously introduces tutors to effective strategies while working to mitigate those influences that lead tutors to eschew the strategies. Thus, the benefits of this line of research are potentially broad. Rather than producing another training program that would compete with existing programs, this research may lead to techniques which could be applied broadly to improve many tutoring programs. This project considered two potential causes of the knowledge-telling bias: inadequate metacognitive self-monitoring and negative motivational attitudes and self-perceptions, both of which are under-researched aspects of tutoring.

The decision to focus on metacognition and motivation should not be construed as a rejection or dismissal of other potential causes of the knowledge-telling bias. For example, another highly plausible cause is peer tutors' perceptions of the tutoring role (Allen & Feldman, 1976; Bierman & Furman, 1981). Tutors' beliefs about what constitutes proper or effective tutoring would almost certainly affect how they choose to teach. For example, some tutors may believe that they should only convey the information in an unmodified form. That is, the source materials contain "the facts," and tutors' job is just to make these facts available to tutees. Generating new examples or analogies might be seen as undesirable because they could confuse the tutees or won't be on the test. Such tutors would likely show a strong knowledge-telling bias.

Little is known about students' beliefs about tutoring, but Foot et al. (1990) described interviews with 8-11 year olds about their beliefs regarding teachers. The children reported that teachers should help others learn facts and skills by explaining, giving demonstrations, and asking questions. Socially supportive behaviors were also mentioned, such as being friendly, patient, or funny. These findings suggest that children perceives both informational ("e.g. give good explanations") and interpersonal (e.g. "be friendly") aspects of the teacher role. However, much more specific data would be needed to determine what kinds of perceptions would lead to a knowledge-telling bias. Such beliefs are also likely to be idiosyncratic and highly variable across students. The current study was not designed to quantitatively assess such beliefs.

Another factor that may be important may be peer tutors' general academic ability. Some tutors may be more intelligent or strategic in their learning behaviors, or simply more "diligent" in the pursuit of learning. These kinds of students might perform well in any setting, whether tutoring a peer, reading a text, writing a term paper, etc. This kind of "good student" quality is difficult to assess in a research study, where one typically has little knowledge of students' performance, aptitudes, or behavioral patterns outside of the experimental context. However, some indicators of this factor can be collected via self-report (e.g. GPA and VSAT scores). Participants' knowledge assessment scores can also be telling. For example, high pre-test scores may indicate students who have a stronger academic background. The current study was not designed to assess participants general academic aptitudes in an in-depth manner, although future analyses may be able identify such a factor through factor analytic methods.

1.2.1 Metacognitive barriers to reflective knowledge-building

Researchers have generally described metacognition as two related component processes, the "monitoring" of cognition and the "regulation" of cognition (e.g. Flavell, 1979; Hacker, 1998). Monitoring refers to an awareness and evaluation of one's cognitive state. For example, imagine a tutor who stumbled over explaining a formula and realized that she did not understand the relationship between two variables. She thus became aware of a gap in her knowledge. Regulation refers what one does in order to change or improve one's cognitive state. Having experienced some confusion, the tutor might have returned to the textbook to find a discussion of the formula, and then generated an example that applied the formula. Thus, the tutor filled a gap in her knowledge and used a knowledge-building strategy to make sure that she understood. Such self-monitoring and regulation are defining aspects of reflective knowledge-building, and a breakdown in either metacognitive process could negate many learning opportunities. Each potential breakdown is discussed in turn.

Not surprisingly, research on metacognition had found that accurate and effective self-monitoring is difficult (e.g. Hacker, 1998). For example, a classic finding in the literature is that learners have problems with successful detection of errors and contradictions (e.g. Baker, 1979; Glenburg, Wilkinson, & Epstein, 1982; Pressley, Ghatala, & Woloshyn, 1990). Learners are sometimes unaware of when one sentence contradicts another sentence in a text, or when their own understanding contradicts the information given in the text. Learners may also not recognize when their own beliefs contradict each other. Ample research on misconceptions and conceptual change shows that students are quite capable of possessing incompatible beliefs without realizing it (e.g. Chi & Roscoe, 2002). Students can also have trouble with accurately judging their level of comprehension or competence (Koriat & Bjork, 2005; Kruger & Dunning, 1999; Maki et al. 2005), typically overestimating how well they have performed or will perform in the future. Learners may attention to inappropriate cues that lead them to falsely believe they understand.

Similar problems likely arise for peer tutors. One problem may be that peer tutors do not monitor their understanding and explanations often enough. If they are confident or overconfident in their perceived expertise, they may simply assume that their explanations make sense and are correct. Another concern is that the quality of tutors' self-monitoring could be poor. That is, even when peer tutors do evaluate themselves, they might fail to detect a problem. Tutors might also pay more attention to trivial matters such as forgetting a key term rather than deeper issues such as agreement between the text and one's mental model. As a consequence of either insufficient or faulty self-monitoring, peer tutors would not detect important knowledge gaps or misconceptions, and these barriers to understanding would persist.

Research has also shown that students have difficulty choosing appropriate strategies to regulate their cognition (e.g. Azevedo and Cromley, 2004; Hacker 1998; Palincsar and Brown 1984; Schraw, 1998). Thus, even when knowledge gaps and confusions are detected, learners may not be able to address these problems effectively. Students' reading strategies tend to focus heavily on memorizing and paraphrasing the text contents, rather than on making predictions, elaborating text ideas with prior knowledge, inferring connections among text concepts, etc. (e.g. Hacker, 1998; McNamara, 2004). Similarly, when learning from examples, students often repeatedly re-read examples rather than trying to infer the purpose of example steps (e.g. Chi et al. 1989). Azevedo and colleagues (e.g. Azevedo & Cromley, 2004; Azevedo, Cromley, & Siebert, 2004; Azevedo, Guthrie, & Siebert, 2004) have collected data on students' monitoring and regulatory behaviors in learning with hypermedia. Students in their control conditions (i.e. were not taught self-regulation strategies) tended to use ineffective strategies such as rereading, searching for information without a specific goal, or simply copying information from the source to their notes. They less often employed strategies that involved elaboration of prior knowledge, inferences, or generating hypotheses and predictions about the material.

Peer tutors may similarly have difficulty selecting strategies that will help them repair knowledge gaps or misconceptions. When a difficult or confusing concept is encountered, some tutors might skim over the topic, thus negating an opportunity to revise their knowledge (and also 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. If peer tutors' metacognitive self-evaluations are ignored, or only addressed through shallow strategies, their learning should be hampered.

The role of metacognition in tutor learning has been neglected (Roscoe & Chi, in-press-B). However, one study (Roscoe & Chi, in-press-A) examined tutors' monitoring behaviors, showing that self-monitoring was linked to higher test scores and knowledge-building. The authors analyzed metacognitive monitoring at a relatively coarse grain-size, classifying segments of verbal explanations as metacognitive if no overt monitoring statements (e.g. "I am confused" or "This is easy to remember") were observed, or as non-metacognitive if one or more monitoring statements occurred. They found that the frequency of metacognitive episodes was significantly and positively correlated with post-test measures of fact recall and comprehension. Non-metacognitive episodes were only weakly or negatively associated to assessment test scores.

These results could be expanded by considering specific features of tutors' metacognitive statements. One variable may be the valence of tutors' judgments, such as positive (This is easy) or negative (I don't understand) statements. Negative statements should be important because they show that the learner has detected an error or comprehension failure, which could then be repaired. Another factor might be the content of the evaluations, such as whether they focus on one's comprehension, ability to remember, correctness, etc. Monitoring comprehension might be especially important for effective learning because it focuses on deeper issues of understanding. Roscoe & Chi (in-press-A) also assessed links between questions and responses. A similar analysis could be used assess how tutors used and responded to their self-evaluations. Such data would allow us to better understand the nature of peer tutors' self-monitoring activities, and determine whether inadequacies of metacognition are a barrier to tutor learning.

1.2.2 Motivational barriers to reflective knowledge-building

It is important not to underestimate the degree of effort and risk-taking that effective tutoring and tutor learning entail. It was argued above that tutors learn better when they confront their own knowledge gaps and use logical reasoning to develop correct explanations. But what does this mean on a personal level for the peer tutor? The tutors will have to admit their lack of knowledge to the tutee, and will then be under pressure to quickly make sense of confusing material and verbalize a correct explanation. This situation has the potential for significant anxiety and embarrassing failure. Thus, peer tutors' beliefs in their ability to engage in these difficult tasks, and their willingness to do so, may influence their decisions about how to tutor. These beliefs are captured in two fundamental motivational variables: self-efficacy and interest.

1.2.2.1 Self-efficacy

Self-efficacy can be defined as "people's beliefs about their capabilities to exercise control over their own level of functioning" (Bandura, 1993). In education settings these perceptions are centered on one's ability to comprehend the material, complete assignments, earn good grades, etc. Studies linking efficacy beliefs and learning have found that students who report higher self-efficacy are more likely to persist after failure, choose more difficult tasks, and exert more effort (Pajares, 1996; Pintrich, 1999; Schunk, 1995; Zimmerman, Bandura, & Martinez-Pons, 1992).

Two types of self-efficacy might be relevant to tutors. First is tutors' self-efficacy regarding their understanding of the subject matter. Tutors who feel capable of making sense of the material may be more willing to engage in knowledge-building activities such as creating novel examples and trying to answer challenging questions. Tutors with a lower sense of "learning efficacy" may be afraid of making mistakes or appearing stupid, and thus elaborate as little as possible and evade difficult questions. Another factor may be tutors' beliefs about their tutoring skills. Tutors who doubt their ability to explain effectively may focus on paraphrasing the textbook and skim over tougher topics or questions. Tutors with higher "tutoring efficacy" may be more willing to take risks and generate new ideas even if they are "just a guess."

1.2.2.2 Interest

Interest can be broadly defined as positive feelings of enjoyment and value of some activity, object, or topic, which reflect a temporary or enduring preference (Deci, 1992; Pintrich, 2003; Schiefele, Krapp, & Winteler, 1992). In learning situations, students' interests reflect the learning domains and tasks that they find stimulating and worthwhile. Similar to self-efficacy, researchers have shown that academic interests influence academic decision-making (e.g. the selection of courses and majors), the effort students put forth, persistence on challenging tasks, and their choice of deeper versus shallower learning strategies (Brophy, 1999; Deci, Ryan, & Williams, 1996; Lepper & Hodell, 1989; Pintrich, 1999).

Two aspects of interest are likely to be relevant to tutor learning. One factor is their interest in the domain. If tutors find the material boring or useless, they may not be willing to thinking deeply about the information. Instead, they may use knowledge-telling strategies to cover the material as quickly as possible, skimming over those ideas they find uninteresting. However, tutors who find the topic intriguing may try to examine and understand ideas more deeply, and try to connect the concepts to their prior knowledge. Interested tutors may also try harder to address knowledge deficits, because they have an intrinsic desire to learn the material. Similarly, tutors may also differ in their interest and enjoyment of the tutoring task. Some students may find teaching a peer to be fun and enjoyable, and thus put more effort and persistence into discussing the material. Those students who dislike being thrust into a teaching role, however, may resist engaging any activity harder than paraphrasing the text. This latter issue may be a particular problem in tutoring programs where whole classrooms or schools participate in the program, and thus tutoring is not on a purely volunteer basis.

Motivational variables in peer tutoring have typically been studied as outcomes of tutoring rather than predictors. Several researchers have hypothesized that tutoring improves tutors' self-efficacy and interest (Cohen, 1986; Benware & Deci, 1984; Fantuzzo, King, & Heller, 1992; Fantuzzo, Riggio, Connelly, & Dimeff, 1989; Gartner et al., 1971), but studies have produced mixed results. Benware & Deci (1984) reported that studying with a teaching expectancy increased enjoyment of the material relative to studying for a test, but Renkl (1995) found a decrease in interest and an increase in anxiety. Similarly, Fantuzzo and colleagues (Fantuzzo et al., 1992; Fantuzzo et al., 1989) observed that reciprocal tutors increased in self-efficacy while decreasing in anxiety, but Rittschof & Griffin (2001) were unable to replicate that finding. Meta-analyses of the effects of tutoring on student affect have consistently found only weak or moderate effects (Cohen et al., 1982; Ginsburg-Block, Rohrbeck, & Fantuzzo, 2006).

Very few studies have examined the impact of motivation on tutors' behavior. Medway & Baron (1977) found that tutors who felt more capable of helping tutees solve problems (i.e.

higher efficacy) exerted more effort towards tutoring. Gabarino (1975) observed that offering tutors a reward based on tutees' performance undermined tutors' interest in tutoring, leading them to be more critical of their tutees, feel less positive about tutoring, and be less effective instructors. Tutors who were not offered a reward reported more enjoyment, were more supportive and patient, and their tutees learned more. Thus, tutors' attitudes and self-perceptions seemed to affect their tutoring behaviors. These results are paralleled by research on "teaching efficacy" (Tschannen-Moran & Hoy, 2001) showing that classroom teachers' sense of efficacy in teaching is associated with more persistence and openness to innovation in the classroom.

1.3 SUMMARY OF HYPOTHESES

The current study address a number of hypotheses regarding tutor learning, knowledge-building, and the knowledge-telling bias. These hypotheses include replications of prior tutor learning research, and novel hypotheses that have not been previously addressed. Each hypothesis is denoted below by a letter in brackets, which appears after the hypothesis has been stated. These notations will also be used in the analyses sections to refer back to the hypotheses.

It is hypothesized that stronger tutor learning outcomes (i.e. higher post-test scores) will be associated with peer tutors' engagement in reflective knowledge-building behaviors [A]. However, peer tutors should display a knowledge-telling bias [B], which will be unrelated or negatively related to learning [C]. This knowledge-telling bias should persist for tutors who are taught knowledge-building strategies, but not given practice and training on these strategies [D]. Hypothesis D is a novel hypothesis.

Tutee questions should have a measurable influence on tutor behaviors. Deep questions should elicit deeper and more metacognitive responses, and shallow questions should elicit shallower and less metacognitive responses [E]. It is further predicted that deep tutee questions, especially those which receive deep answers, should be linked to stronger learning outcomes, whereas shallow questions and responses should not support learning [F]. Hypothesis F extends prior work, which has not shown direct connections between question patterns and learning.

Metacognitive monitoring should be an important aspect of knowledge-building and tutor learning, and should be related to stronger learning outcomes [G]. However, there may be differences based on whether the statement relates to comprehension-monitoring, memory-monitoring, accuracy-monitoring, etc. Although strong a priori predictions are not given, comprehension-monitoring may be most critical [H]. Similarly, there may be differences based on the valence of the monitoring statements. Because negatively-valenced statements represent detection of knowledge-gaps and errors, these may be particularly salient locations for tutor learning [I, novel]. Finally, tutors' utilization of these self-judgments should be related to learning outcome. Tutors whose metacognitive statements occur in conjunction with elaborations, representing attempts to elaborate or repair their knowledge, should learn more effectively [J]. Hypotheses H through J are novel hypotheses.

Several hypotheses regarding motivation are also considered. Tutors who express higher learning efficacy [K], tutoring efficacy [L], topic interest [M], and/or tutoring interest [N] should be more likely to engage in reflective knowledge-building and/or less likely to engage in knowledge-telling. As a result, more self-efficacious and interested peer tutors should also show stronger learning outcomes. No predictions are made about which, if any, of these motivational variables will have a greater impact on learning. Hypotheses K through M are novel hypotheses.

2.0 METHOD

2.1 PARTICIPANTS

Participants were 140 college undergraduates from the University of Pittsburgh, who received course credit in an Introductory Psychology class or cash payment for their time. Participants filled out a Background Questionnaire (Appendix A) that asked them to report demographic (e.g. age and gender) and academic information (e.g. GPA and verbal SAT). The questionnaire also asked about prior experiences with tutoring. The primary concern was students' formal tutoring experiences (i.e. supervised tutoring in an established program).

Table 1 summarizes age, gender, ethnicity, year in school, verbal SAT scores, and tutoring experience. Many students claimed to forget their specific SAT scores, and so these data were of dubious quality. GPA scores were omitted because a many participants were first-semester freshmen, and thus had no college GPA to report. Non-significant chi-square tests showed that male and female students were distributed evenly across groups with slightly more female participants than male participants. Students were primarily Caucasian, although there were some African-American, Asian, and Hispanic participants, who were distributed fairly evenly across conditions. Two students chose not to report their ethnicity. Few students in the study had any formal tutoring experience, and these students were relatively evenly distributed. There was no difference in whether participants were science or non-science majors.

Table 1. Distribution of demographic factors.

Variable	Groups: Reviewers n = 20	Tutors n = 60	Tutees n = 60	Overall mean/total
Mean age	21.2 (2.5)	19.2 (1.4)	19.4 (1.92)	19.6 (1.9)
Mean VSAT	637.8 (81.6)	619.7 (72.0)	625.8 (77.4)	624.6 (75.1)
Gender				
Female	11	33	33	77
Male	9	27	27	63
Ethnicity				
Caucasian	14	42	42	98
African-Amer.	3	9	10	22
Asian	2	7	6	15
Hispanic	1	1	1	3
Year				
1st year	2	30	30	62
2nd-4th year	17	19	20	56
Major				
Non-science	13	36	32	81
Science	7	24	28	59
Tutoring				
No formal exp.	14	48	45	107
Formal exp.	6	12	15	33

Note. Data for age and VSAT scores are means. Data for all over variables are frequency counts.

There was a significant difference in age, $F(2, 137) = 9.22$, $p < .001$. Reviewers were about two years older than tutors and tutees, on average, and so thus were also more upperclassmen in the reading group than in the tutor and tutee groups, $\chi^2(2, N = 118) = 16.0$, $p < .001$. 22 students did not report their year in school. There was no difference in reported verbal SAT scores. 28 students failed to report a verbal SAT score.

2.2 MATERIALS

2.2.1 Human eye and retina text

The domain was the structure and function of the human eye and retina. Some prior studies on tutor learning (e.g. King et al., 1998; Roscoe & Chi, in-press-A) have used physiology domains, and so this choice offers comparability to such work. The text described the regulation and focusing of light via the cornea, iris, lens, and related structures (Appendix A). The text also described the function and distribution of photoreceptors, as well as several retinal neurons (e.g. bipolar and ganglion cells). The text indicated that neural signals travel to the brain via the optic nerve, but did not describe any visual processes that occur in the brain. Two labeled diagrams were provided: a cross-section diagram 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 cornea and iris) and unfamiliar, technical concepts (e.g. neural connections of bipolar cells). This provided ample opportunities to make connections with prior knowledge and explore new ideas.

This material was informed by college textbooks on sensation and perception (Goldstein, 1999) and human physiology (e.g. Fox, 1996). The information was reduced to the level of basic definitions of components and short descriptions of their functions. No examples and analogies were included, nor were advanced references to biochemistry (e.g. synaptic transmission), optics (e.g. inversion of light by the lens), or disorders of the eye (e.g. glaucoma). The resulting text was 59 sentences in length (848 words), divided into 18 topic-based paragraphs presented on separate pages without headings. Readability was assessed using tools available through Coh-Metrix (McNamara, Louwerse, Cai, & Graesser, 2005), which indicated a Flesch Reading Ease score of 57.5 and a Flesch-Kincaid Grade Level score of 8.8.

2.2.2 Knowledge assessments

Learning outcomes were assessed using two written measures (Appendix A). For the Definitions Test, students labeled blank versions of the eye and retina diagrams (target components were indicated by arrows) and defined each labeled component. Students were encouraged to be accurate and thorough. There were 11 targets on the eye diagram, and 6 on the retina diagram. Participants were given a list of terms to use, and all of terms were to be used. 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. For example, one question asked about 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 information 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 text information. This test can be considered a measure of participants' deeper learning.

The core test consisted of 4 questions covering topics of focusing of light, neural pathways, and photoreceptor functions and distribution. This test was included in all learning assessments throughout the study. On the post-test, 6 more questions were added that covered additional topics. The 4 "core" questions occurred at the beginning of the post-test, followed by the 6 "new" questions. The purpose of the new questions was to test learners' ability to answer questions they had not previously seen. These questions are labeled in Appendix A.

2.2.3 Self-efficacy and interest scales

Self-efficacy and interest were assessed using scales in which participants rated their agreement with statements about perceived abilities and interests. These scales were inspired by prior measures used to assess students' and teachers' attitudes in classrooms (i.e. Motivated Strategies for Learning Questionnaire, Pintrich & De Groot, 1990; Ohio State Teaching Efficacy Scale, Tschannen-Moran & Hoy, 2001). These measures guided the types of statements included in the measures. For example, measures of learning efficacy might include statements about one's ability to comprehend the material, knowledge of good strategies, and favorable comparisons to others. Measures of interest might include statements about the utility, value, and enjoyment of the material. Several researchers (e.g. Pajares, 1996; Schiefele et al. 1992) have argued that efficacy and interest beliefs are often specific to a topic or task. Thus, items were worded to relate to the specific study, rather than general learning or tutoring situations.

Two scales were developed to assess participants' attitudes (Appendix A). The Attitudes Toward the Material Scale (ATMS) consisted of 8 items to assess students' self-efficacy for learning the material or "learning efficacy," and 8 items to assess students' feelings of enjoyment and value of the material or "topic interest." The Attitudes Toward Tutoring Scale (ATTS) included 8 items to assess students' self-efficacy for teaching the material effectively or "tutoring efficacy," and 8 items to assess students' enjoyment and value of the tutoring task or "tutoring interest." Each item was one sentence in length and participants rated their agreement using a 7-point Likert-like scale. On this scale, a "1" represented "strongly disagree" and a "7" represented "strongly agree." Half of the items were positively worded, such that agreement meant higher efficacy or interest. The remaining items were negatively worded, such that agreement indicated lower efficacy or interest. Students were instructed that there were no right or wrong answers, but to read each item carefully and respond as honestly as possible.

These scales were intended to yield four scores: learning efficacy, tutoring efficacy, topic interest, and tutoring interest. A principal components analysis with varimax rotation was conducted to confirm this factor structure. The data for this analysis came from the first administration of the scales during the pre-test, which was completed by all 140 participants. A separate analysis was conducted for each scale.

The Attitudes Toward the Material Scale produced five components with eigenvalues greater than 1, which accounted for 67.4% of the variance (Table 2). Using a threshold of .30 for factor loadings, the clustering of items within these components was consistent with planned scales. Learning efficacy items loaded most strongly on components 1, 4 and 5. Topic interest items loaded most strongly on components 2 and 3. There was little overlap. The Attitudes Toward Tutoring Scale produced three components with eigenvalues greater than 1, which accounted for 57.0% of the variance (Table 3). Tutoring efficacy items loaded most strongly on components 2 and 3. Tutoring interest items loaded most strongly on component 1, and also on component 3 to a lesser extent. There was little overlap overall. In sum, these analyses suggest that the scales can be separated into the four desired constructs.

It is important to note that these constructs, although separable, were related to one another. This makes sense, because one might expect learning efficacy beliefs to influence one's judgments about tutoring abilities. Similarly, enjoyment of the tutoring task might lead learners to also feel more positively about the material being covered. Efficacy and interest beliefs should also be related (Deci, Ryan, & Williams, 1996). For example, feeling incompetent or unable to learn the material may lead to reduced interest and enjoyment. These relationships are captured in Table 4, which reports intercorrelations between scale scores for all participants ($N = 140$).

Table 2. Components and factor loadings for ATM scale

	Components and factor loadings				
Item	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5
LE Item 1	.719				
LE Item 2				.838	
LE Item 3					.858
LE Item 4	.711			.351	
LE Item 5	.838				
LE Item 6				.584	.521
LE Item 7	.407			.480	.472
LE Item 8	.793				
TI Item 1		.822			
TI Item 2		.342		.596	
TI Item 3		.595	.508		
TI Item 4		.592	.426		
TI Item 5			.769		
TI Item 6			.653		
TI Item 7		.766			
TI Item 8			.804		

Note. LE = “Learning Efficacy” and TI = “Tutoring Interest”

Table 3. Components and factor loadings for ATT scale

	Components and factor loadings		
Item	Comp 1	Comp 2	Comp 3
TE Item 1		.707	
TE Item 2			.684
TE Item 3	.400	.662	
TE Item 4		.700	.314
TE Item 5		.740	
TE Item 6		.468	.592
TE Item 7		.571	
TE Item 8			.637
TU Item 1	.775		
TU Item 2	.717		
TU Item 3	.792		
TU Item 4	.380		
TU Item 5	.820		.531
TU Item 6			
TU Item 7	.790		.681
TU Item 8	.478		

Table 4. Interrcorrelations for motivation scale scores.

	Motivation scales		
	Topic Interest	Tutoring Efficacy	Tutoring Interest
Learning efficacy	.43[a]	.45[a]	.20[c]
Topic interest		.35[a]	.54[a]
Tutoring efficacy			.55[a]

[a]$p < .001$, [b]$p < .01$, [c]$p < .05$

2.3 CONDITIONS AND PROCEDURES

At the beginning of the study participants in the study were randomly assigned to one of three main conditions: peer tutoring (as the tutor), peer tutoring (as the tutee), or active reviewing. Many more students were recruited for the tutoring condition because tutor learning and behavior were the focus of the study, and tutors were divided into three subgroups based on differing instructions. The purpose of the reviewing condition was primarily to provide baseline learning data. Participants completed the study in a single session, about 2 hours in length, which was divided into five phases: pre-test, studying, mid-test test, tutoring/reviewing, and post-test. These phases are summarized graphically in Table 5.

Table 5. Summary of phases, tasks, and assessments

<table>
<tr><th></th><th colspan="5">Phases of the Study</th></tr>
<tr><th>Condition</th><th>Pre-test</th><th>Study</th><th>Mid-test</th><th>Review/Tutor</th><th>Post-test</th></tr>
<tr><td>Reviewers</td><td rowspan="3">questionnaire
definitions
4 questions
motivation</td><td rowspan="2">20 minutes to study text and take notes</td><td rowspan="2">definitions
4 questions
motivation</td><td>20 min. for think-aloud</td><td rowspan="3">definitions
4 questions
+ 6 questions</td></tr>
<tr><td>Tutors</td><td>20 min. for peer tutoring</td></tr>
<tr><td>Tutees</td><td colspan="2">waiting</td><td>tutored</td></tr>
</table>

In the pre-test phase, all participants (i.e. reviewers, tutors, and tutees) completed the Background Questionnaire, Definitions Test, Questions Test (4 questions), ATMS, and ATTS. Knowledge assessments were completed before motivation assessments. Participants received more specific instructions (see below) immediately after the pre-test.

In the studying phase, peer tutors and reviewers were given 20 minutes to study the text and diagrams. Participants were encouraged to take notes, and were told that their notes and the text would be available during the later tutoring or reviewing phase. Afterwards, tutors and reviewers completed the mid-test assessments, which were identical to the pre-test assessments. During the studying and mid-test phases, the tutees were escorted to a separate waiting area. The tutees were given a blank sheet of paper and encouraged to generate questions that they might like to ask the tutor. No specific guidelines were given regarding the creation of the questions. Tutees did not read the text or take the mid-test assessments.

In the tutoring/reviewing phase, peer tutors were given 20 minutes to teach the material to their tutee using their assigned instructions. Tutors were allowed to use the text, diagrams, and their notes. They were not permitted to simply read aloud from the text to the tutee, and tutees could not read the text on their own. These rules were instituted to ensure that the tutors had to actually explain the material to the tutee. Tutees were allowed to take their own notes, however, and had free access to the diagrams. During this phase, the reviewers were given 20 minutes to review their text, diagrams, and notes using a think-aloud strategy. All of the tutoring/reviewing sessions were videotaped. Afterwards, all participants (i.e. reviewers, tutors, and tutees) completed the post-test assessments. These assessments included the Definitions Test and Questions Test (4 core questions plus 6 new questions), but not the motivation scales.

2.3.1 Peer tutoring instructions

One hypothesis is that tutors do not use effective strategies because they are unaware of them. Student tutors are probably often exposed to knowledge-telling lectures, but not other methods. Thus, tutors were divided into three groups that varied in the tutoring instructions they received (Appendix B). All tutors received a card summarizing their strategies. In addition, all tutors were encouraged to try to improve their answers and their tutees' answers to the Definitions Test and Questions Test, and were reminded that the Questions Test contained challenging comprehension questions. The three groups did not differ on any of the demographic variables.

General tutors (n = 20) received only basic instructions to cover all of the material in the text, give good explanations, and to try to respond to their tutees' questions. They were not given any specific explaining or question-answering strategies.

Explanation tutors (n = 20) were introduced to three explanation strategies, which were based on tutoring behaviors found to be beneficial in Roscoe & Chi (in-press-A). Tutors were instructed to give detailed and elaborated definitions that contained new examples and elaborations, and explained how major concepts are related. Tutors were also instructed to give elaborated reviews of concepts, which did not simply reiterate previous ideas but contained new elaborations and sought to fit concepts together into a "big picture." Finally, tutors were reminded that they would sometimes be unable to explain an idea because they did not understand. When this occurred, tutors were encouraged not to skip over the concept, but to use all available resources to try to make sense of the ideas and work out a correct explanation.

Question tutors (n = 20) were introduced to the same three strategies as the explanation tutors, except that they were described as question-answering strategies. When answering questions, tutors were instructed to include elaborated definitions with new examples and connections. When asked review questions, tutors were told not to only repeat prior ideas, but to offer new elaborations. Finally, tutors were reminded that they would sometimes be unable to answer questions due a lack of understanding. Tutors were told to not ignore such questions, but to use available resources to make sense of the ideas and work out a correct answer.

It should be stressed that these conditions did not involve tutor training. Unlike training studies, tutors were not given days of instruction and practice. However, tutors who wished to use deeper strategies had such strategies available. Thus, a lack of awareness of knowledge-building strategies might be evaluated as an explanation for the knowledge-telling bias.

2.3.2 Active reviewing instructions

Active reviewers (n = 20) were instructed to use a form of "think-aloud" strategy during the reviewing phase of the study (Appendix B). Active reviewers were instructed to say out loud everything that they were thinking and feeling about the material, to think out loud at least once or more for each major concept, and to continue to think out loud for the duration of the reviewing phase. An experimenter was not present to prompt the reviewers. The reviewers were encouraged to try to improve their answers on both the Definitions Test and Questions Test, and were reminded that the Questions Test contained challenging comprehension questions.

This kind of think-aloud type of reviewing activity was much more active than students' typical reading and studying of text information. Although the active reviewers were not explicitly instructed or prompted to self-explain, the nature of the task could induce some students to spontaneously self-explain (Chi, 2000). As a result, the active reviewing condition provided a fairly challenging comparison condition by which to contrast tutors' learning.

3.0 ANALYSES AND RESULTS: KNOWLEDGE ASSESSMENTS

3.1 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. The Questions Test was scored by tabulating the number of correct responses students gave in their answers. Correct responses were defined as any accurate, relevant, and unique statement that addressed the core question being asked. Irrelevant information was ignored. This scoring allowed participants to earn partial credit for incomplete responses, but participants could not inflate their score by repeating the same statements across multiple questions. To reduce ambiguity in scoring, an answer template was generated and students' responses were compared to these templates. The templates were modifiable in the rare case that students produced an unanticipated but plausible answer.

Neither the Definitions Test nor Questions Test had an absolute maximum score. Table 6 summarizes the minimum and maximum scores obtained by any participant on the post-test administrations of these tests, along with the grand means.

Table 6. Descriptive statistics for post-test assessment scores

Measure	Min	Max	Mean	SD
Definitions Test	4.0	49.0	23.0	8.8
Questions Test	3.0	53.0	21.3	10.3
4 Core questions	0.0	31.0	10.3	5.9
6 New questions	1.0	26.0	11.0	5.2

3.2 PERFORMANCE OF THE TUTEES

The primary focus was on tutor learning, but it was critical to show that the tutors were effective instructors. Definitions Test gains were assessed in a 2 (test) x 3 (condition) mixed-factor repeated measures ANOVA. Tutees gained significantly from pre-test to post-test, $F(1,57) = 386.77$, $p < .001$, $R^2 = .87$. A similar analysis showed that tutees gained significantly on the 4 Questions Test core questions from pre-test to post-test, $F(1,57) = 172.39$, $p < .001$, $R^2 = .75$. There were no effects of tutoring condition or interactions for either test. Tutees' scores on the 6 new questions and their total Questions Test scores also did not differ by condition (Table 7).

In sum, tutees made measurable gains in their factual recall and comprehension of eye concepts. These gains were impressive when one considered that the tutors were only given 20 minutes to initially learn the material and 20 minutes to teach it. The tutees also never read the text. However, tutee learning was not affected by the strategy instruction received by the tutors. One possible explanation is that the strategies believed to benefit tutor learning did not affect tutee learning. However, it is more likely that the tutors did not actually differ in their instructional behaviors, and thus the tutoring subgroups were essentially homogeneous.

Table 7. Mean assessment test scores for tutees

	Condition		
Measure	General tutees (n = 20)	Explain tutees (n = 20)	Question tutees (n = 20)
Definitions Test			
Pre-test	2.7 (2.9)	4.3 (2.5)	3.3 (3.1)
Post-test	18.4 (6.3)	23.2 (6.1)	20.7 (8.2)
Questions Test			
Pre-test (4 core)	1.1 (1.4)	1.4 (1.3)	1.4 (1.5)
Post-test (4 core)	7.6 (4.6)	10.5 (5.2)	8.4 (4.3)
Post-test (6 new)	8.2 (4.1)	10.4 (4.8)	9.2 (3.5)
Post-test (total)	15.8 (7.6)	20.8 (9.2)	17.6 (7.1)

Another important question was whether tutee learning was connected to or dependent upon tutor learning. One might imagine that tutors who better understood the material would be better able to teach it. They had more information that could be shared with the tutee. There are two alternatives, however. One might find no relationship between tutor and tutee test scores, implying that tutor learning and tutee learning were independent of each other. For example, a good tutee might be able to learn effectively even with a less knowledgeable tutor. Another alternative is that tutor and tutee learning might be negatively related to one another. If peer tutors focus too much on their own understanding, the needs of the tutee might be neglected.

One might expect that peer tutors who learned more during the initial studying phase would have more knowledge to draw from when teaching their tutee. Because the focus is on tutor learning during the tutoring session, this prior knowledge was controlled for by entering tutors' mid-test Questions Test scores (on the 4 core questions) as a covariate. Partial correlations were computed between tutor and tutee post-test scores, as well as tutee gain scores (post-test minus pre-test scores). Table 8 summarizes these partial correlations. It is evident that tutor and tutee scores were positively correlated in all cases, and significantly correlated in many cases. Thus, there did not seem to be a tradeoff between tutor and tutee learning. Indeed, it seems possible that tutor and tutee learning were mutually supportive. As tutors improved their understanding, they could perhaps share these insights and corrections with tutees. As tutees learned, they might have engaged the tutor in deeper conversations about the material.

Table 8. Partial correlations between tutor and tutee post-test scores

	Peer Tutor Post-test Scores			
Peer Tutee Post-test Scores	Definitions Test	Questions Test (4 core)	Questions Test (6 new)	Questions Test (total)
Definitions Test				
Post-test total	.35[b]	.36[b]	.33[b]	.39[b]
Pre-post gain	.36[b]	.34[b]	.32[c]	.38[b]
Questions Test				
4 core questions	.17	.42[a]	.28[c]	.41[a]
Pre-post gain	.17	.43[a]	.29[c]	.42[a]
6 new questions	.14	.32[c]	.19	.30[c]
Total questions	.17	.42[a]	.27[c]	.40[b]

[a]$p \leq .001$, [b]$p < .01$, [c]$p < .05$

3.3 PERFORMANCE OF THE TUTORS AND REVIEWERS

Tutor and reviewer learning were assessed by conducting a 3 (time) x 4 (condition) mixed-factor repeated measures ANOVA. Participants made significant gains on the Definitions Test from pre-test to mid-test, $F(1,76) = 238.30$, $p < .001$, $R^2 = .76$; and from mid-test to post-test, $F(1,76)$

= 95.41, $p < .001$., $R^2 = .55$. There was no interaction with condition. Participants across conditions gained similarly in factual knowledge (Table 9).

Gains on the 4 Questions Test core questions were similarly assessed. Participants gained significantly from pre-test to mid-test, $F(1,76) = 108.21$, $p < .001$, $R^2 = .59$; and from mid-test to post-test, $F(1,76) = 56.44$, $p < .001$, $R^2 = .43$. Again, there was no significant effect or interaction with condition. One-way ANOVAs were used to compare performance on the 6 new post-test questions and overall post-test scores. No significant differences were found between conditions. Participants across conditions improved similarly in their deeper understanding of the material.

Overall, gains from pre-test to mid-test showed that students benefited from their initial studying of the text. Gains from mid-test to post-test showed that students also learned from their think-aloud or tutoring activities. Thus, tutors did learn while simultaneously tutoring a peer. However, by a strong definition of tutor learning the tutor learning effect was not observed because tutors did not surpass reviewers. This is not uncommon (Roscoe & Chi, in-press-B). However, it should be remembered that the think-aloud instructions may have lead some students to self-explain, which could supported stronger learning outcomes (Roscoe & Chi, in-press-A).

There was also little evidence that providing tutors with explaining or questioning strategies improved learning outcomes. One possibility is that the chosen strategies were not effective. This is unlikely because these strategies were derived directly from behaviors shown to be effective in prior research (Roscoe & Chi, in-press-A). Another possibility is that tutors ignored these knowledge-building strategies and focused on knowledge-telling. In the next sections, a series of analyses are presented examining peer tutor's behaviors, and how the occurrence of these behaviors differed across high-performing and low-performing participants.

Table 9. Mean assessment test scores for tutors and reviewers

	Condition			
Measure	Active Reviewers	General tutors	Explain tutors	Question tutors
Definitions Test				
Pre-test	2.6 (2.2)	2.8 (3.0)	3.4 (2.3)	3.4 (2.5)
Mid-test	17.0 (9.3)	17.4 (9.6)	18.6 (10.3)	17.1 (7.2)
Post-test	24.0 (10.2)	23.9 (9.9)	26.2 (9.8)	24.4 (8.7)
Questions Test				
Pre-test (4 core)	1.4 (1.7)	1.2 (1.6)	1.6 (1.7)	1.2 (1.4)
Mid-test (4 core)	6.8 (6.1)	7.4 (6.4)	9.4 (6.9)	6.5 (4.6)
Post-test (4 core)	10.4 (7.0)	11.0 (6.3)	12.8 (7.3)	11.4 (5.2)
Post-test (6 new)	12.5 (5.3)	10.8 (6.0)	12.8 (5.8)	12.9 (5.3)
Post-test (total)	22.9 (10.9)	21.9 (12.0)	25.5 (12.4)	24.2 (9.5)

4.0 ANALYSES AND RESULTS: LEARNING BEHAVIOR EPISODES

Successful learning should be associated with reflective knowledge-building strategies that integrate new and prior knowledge, generate ideas, and repair confusions. To test this hypothesis, a subset of verbal tutoring and active reviewing protocols were analyzed to identify episodes of knowledge-building and knowledge-telling. In previous research (Roscoe & Chi, in-press-A), analyses at this grain-size were able to capture meaningful instances of learning behaviors.

4.1 DATA SELECTION AND REDUCTION

Transcription of verbal protocols is a time-consuming and expensive process. Due to limitations of time, and especially financial resources, it was necessary to reduce the amount of protocol data to be processed and analyzed. Such data reduction techniques are commonly used in peer tutoring research, however, when tutors' behaviors are analyzed (e.g. Fuchs et al., 1997; Roscoe & Chi, in-press-B). In this study, students' final understanding of the material was a key outcome measure, and so the data were reduced to create two groups of high and low performers based on post-test scores. Differences in the behaviors of these more and less successful learners can provide insights about which behaviors may have enhanced or hindered learning.

Participants within each condition were ranked according to their post-test Questions Test and Definitions Test scores. The six participants ranked 1st through 6th in each condition were identified as "high performers." The six participants ranked 13^{th} through 18^{th} in each condition were identified as "low performers." It was decided to exclude the very worst performers (i.e. ranked 19^{th} and 20^{th}) in each condition. These students' total Questions Test post-test scores were all in the single digits and almost two standard deviations below the mean. These suspiciously low scores may have indicated students who tried and failed in the tasks of the study. However, it was likely that external factors beyond our control had some influence (i.e. illness, anxieties about relationships, time constraints, etc.), resulting in students being too distracted to participate fully in the study. Because any data reduction technique also reduces statistical power, it was decided to try to limit the impact of these potential concerns for the current analyses.

To validate the "high" versus "low" designation, post-test scores were examined in 2 (performance) x 4 (condition) analyses of variance (ANOVA) and covariance (ANCOVA). Pre-test scores were entered as a covariate when comparing mid-test scores, and mid-test scores were entered as a covariate when comparing post-test scores. This method enabled sensitivity to effects resulting from specific learning phases versus prior knowledge (Table 10).

High and low performers differed significantly in their pre-test knowledge of definitions, $F(1,40) = 4.22$, $p = .046$, $R^2 = .10$. However, these scores were very low and the gap between them was not meaningful. A larger difference in factual recall favoring high-performers was found after initially studying the text, $F(1,39) = 60.59$, $p < .001$, $R^2 = .61$, but there was no significant difference in post-test Definitions Test scores There were no condition effects or interactions for any of the Definitions Test assessments.

Table 10. Mean assessment test scores for high and low performers by condition

	Conditions			
	High Performers			
Episode Category	Active Reviewers	General tutors	Explain tutors	Question tutors
Definitions Test				
Pre-test	3.5 (2.2)	4.6 (4.8)	4.8 (1.7)	4.5 (3.2)
Mid-test	26.7 (5.4)	27.1 (5.5)	26.1 (5.5)	21.0 (5.4)
Post-test	28.5 (7.3)	28.5 (7.0)	26.6 (7.4)	28.5 (6.6)
Questions Test				
Pre-test (4 core)	1.5 (2.2)	2.0 (2.1)	2.7 (2.0)	1.5 (1.4)
Mid-test (4 core)	11.9 (4.1)	11.7 (4.2)	15.0 (4.3)	8.6 (4.1)
Post-test (4 core)	17.3 (3.4)	16.8 (3.5)	17.8 (4.0)	17.0 (3.3)
Post-test (6 new)	16.0 (3.2)	14.9 (3.3)	16.5 (3.8)	17.0 (3.1)
Post-test (total)	33.2 (4.2)	32.2 (4.2)	34.5 (4.9)	34.0 (4.0)
	Low Performers			
	Active Reviewers	General tutors	Explain tutors	Question tutors
Definitions Test				
Pre-test	2.8 (2.5)	1.3 (1.0)	3.5 (2.9)	3.2 (2.3)
Mid-test	12.8 (5.4)	12.2 (5.6)	12.2 (5.4)	12.6 (5.4)
Post-test	21.1 (7.0)	23.4 (7.4)	26.9 (7.0)	23.3 (7.0)
Questions Test				
Pre-test (4 core)	1.2 (0.8)	0.3 (0.8)	1.2 (1.6)	1.3 (2.0)
Mid-test (4 core)	3.0 (4.1)	4.5 (4.3)	5.1 (4.1)	5.7 (4.1)
Post-test (4 core)	7.1 (3.6)	7.7 (3.6)	9.9 (3.4)	8.7 (3.4)
Post-test (6 new)	11.8 (3.4)	7.8 (3.4)	10.3 (3.2)	10.0 (3.2)
Post-test (total)	18.8 (4.4)	15.4 (4.4)	20.2 (4.2)	18.6 (4.1)

Note. Mid-test and post-test scores were analyzed using ANCOVA procedures, and thus the means for these assessments are reported as adjusted means.

There was not a significant difference in high and low performers' pre-test Questions Test scores. After reading the text, high performers scored significantly higher on the Questions Test than low performers, $F(1,39) = 33.68$, $p < .001$, $R^2 = .46$. On the post-test, high performers scored higher than low performers on the 4 core questions, $F(1,39) = 44.56$, $p < .001$, $R^2 = .53$; 6 new questions, $F(1,39) = 24.14$, $p < .001$, $R^2 = .38$; and overall, $F(1,39) = 87.93$, $p < .001$, $R^2 = .69$. There were no effects of condition or interactions on any of the Questions Test measures.

In sum, high and low performing learners differed measurably in their factual recall and deeper understanding of the material. Results suggest that high performers' factual knowledge was acquired mainly during the studying phase. One possibility was that the high performers may have been more skilled readers, who were better able to quickly grasp basic facts. Although the verbal SAT score data collected in this study was incomplete (and potentially untrustworthy), an analysis was conducted to compare high and low performers' VSAT scores. There was no difference across experimental conditions, but high performers ($M = 653.4$, $SD = 71.6$) did self-report somewhat higher mean VSAT scores than low performers ($M = 608.3$, $SD = 73.6$), $F(1,30) = 3.61$, $p = .069$, $R^2 = .11$. High performers may have been slightly better at learning from the text. In contrast to Definitions Test scores, differences in Questions Test scores stemmed from both the initial studying phase and the tutoring/reviewing phase. Thus, high and low performers' learning behaviors in that phase may have further improved their deeper comprehension above and beyond what they had already learned.

There was no effect of condition on learning outcomes. Participants within a given level of performance did not seem greatly affected by whether they used a think-aloud strategy or taught a peer, or by what strategies they were instructed to use. This result implies that participants within a given level of performance may not have differed in their learning behaviors. However, high performers likely engaged in more knowledge-building strategies than low performers. These issues are addressed in the following sections.

4.2 CODING OF EPISODES

The verbal protocols of participants were fully transcribed and segmented according to changes in the topic of discussion, which formed the boundaries of "episodes." Thus, each episode was a brief segment of the overall protocol that was devoted to one particular topic. These topics could refer to a single eye component (e.g. rod receptors) or multiple components linked by a single concept or process (e.g. how rod and cone receptors are involved in night vision). An average of 25 episodes were produced per participant (SD = 5.8) in sessions that were a mean duration of 19 minutes (SD = 3.5). Session duration and number of episodes did not differ significantly across high and low performers. Number of episodes did not differ significantly across conditions, but tutoring sessions were about 3 minutes longer than think-aloud sessions on average, $F(1,40) = 3.42$, $p = .026$, $R^2 = .20$. The tutoring sessions involved two speakers who could interrupt each other, thus taking slightly longer to cover the material.

The episodes were characterized by the overall pattern of statements uttered. For example, paraphrase statements simply reiterated text contents verbatim, or with some rewording and clarification. Elaboration statements contained meaningful information that was not given in the text, and thus had to be integrated from prior knowledge or newly generated. Monitoring statements contained self-evaluations of the speaker's comprehension, accuracy, uncertainty, etc. Based on patterns of these statements, five main episode types were identified in our data:

summary, review, concept elaboration, elaborated review, and sense-making. A sixth episode type, passive feedback, occurred only in tutoring sessions.

All of these episodes fit on a continuum between purely knowledge-telling, analogous to exclusively paraphrasing, and knowledge-building, which is analogous to self-explaining. Each episode category is further described below and their relationship to knowledge-telling and knowledge-building are discussed. Coded examples from tutoring protocols are provided, which have been slightly edited for clarity (e.g. stutters were removed and punctuation added).

Reliability of the episode coding process was established by having two coders (the author and a graduate student) code the same randomly selected subset of ten complete protocols. The second coder was presented with a template describing the features of each episode type, along with examples, which were briefly explained verbally. Agreement between coders was reasonable (80.4%) given the challenging nature of the coding.

Table 11 provides the raw mean frequencies of each episode category across groups, which show how much of each activity occurred. The actual analyses of behaviors across high and performers and conditions used proportions (i.e. episode frequency divided by total number of episodes). The use of proportions helps to control for individual differences in how many episodes (i.e. amount of discussion) were produced by individual participants, and highlights important relative differences in the occurrence of episodes. Summary episodes

4.2.1 Summary episodes

In summary episodes, participants primarily paraphrased the text. In some cases, participants created very shallow mnemonic devices based on rhyming words, matching letters, or similar cues. Minor elaborations of non-text information were also sometimes given in these episodes, but there was little attempt to further develop these ideas or integrate them with other concepts. Overall, summary episodes were canonical examples of knowledge-telling, in which very little was discussed beyond the given contents of the text.

Excerpt A provides an example of summarizing in which a low-performing explanation tutor described the iris and the pupil.

[A] Tutor: And then it goes in between these two irises which are right here. (paraphrase)
Tutor: Which are in front of the pupil – which are behind the pupil. (paraphrase)
Tutor: Like, the pupil is in front of the lens. (paraphrase)
Tutor: The iris can open and closes to adjust how much light is entering. (paraphrase)
Tutor: And the iris deals with color. (paraphrase)
Pupil: Okay
Tutor: So the iris's main thing is color. (paraphrase)
Tutor: And then it goes through — light goes in the pupil, through the pupil. (paraphrase)
Tutor: Which I didn't really think it explained too much about the pupil. (comment)
Tutor: It just goes through the pupil. (paraphrase)
Pupil: It's kinda like the hole that the light goes through?
Tutor: Yeah.

4.2.2 Review episodes

In review episodes, participants returned to a previous topic for further discussion. These episodes primarily involved rehearsing or retelling the same information that had been given before, thus repeating prior paraphrases or inferences. In some cases, minor elaborations were added, but with no further development or integration of the ideas. Overall, review episodes were also clearly examples of knowledge-telling.

Excerpt B provides an example of reviewing in which a low-performing question tutor responded to a tutee's review question about the lenses and accommodation.

[B] Pupil: And the lens? What do the lenses do?

Tutor: The lens is a bunch of multiple layers that connect to these ciliary muscles. (paraphrase)

Pupil: Right.

Tutor: The ciliary muscles control the thickness of the lens. (paraphrase)

Tutor: The lens – it doesn't say what it does. (paraphrase)

Tutor: I guess it – Yeah. (monitoring)

Tutor: And our process was called accommodation when the muscles contract. (paraphrase)

Pupil: It's called accommodation?

Tutor: Yeah.

4.2.3 Passive feedback

In passive feedback episodes, the tutee tried to summarize or review the material, while the tutor offered confirmations or minor corrections. Sometimes these episodes were prompted by tutors quizzing their tutees. In these episodes, tutors made very few substantive contributions (i.e. new ideas or elaborations) in these episodes. Thus, passive feedback episodes represented a passive knowledge-telling activity for the tutor (although active for the tutees) because tutors did not advance or improve their own ideas. Passive feedback could only occur in tutoring sessions.

Excerpt C provides an example of passive feedback in which a low performing question tutor quizzed his tutee about the path of light the eye and relevant eye components. The tutors' contributions were limited to confirmations (e.g. "Right" or "Correct"), and a brief clarification in which he corrected himself.

[C] Tutor: Okay. So let's go over these real quick so you get those all right. (comment)

Tutor: First thing – okay, so first it hits the cornea. (paraphrase)

Pupil: And then it goes to the pupil and the pupil contracts or gets smaller depending on the amount of light and it's between the iris. The lens gets thicker or thinner depending on the ciliary muscles and it's like the protective covering of the eye. Like-

Tutor: Well- Yeah. That's – the lens – okay. (confirmation)

Tutor: This is more – I think I was kind of wrong in saying that. (monitoring)

Tutor: Think of like the cornea as more of the protective covering. Okay. (paraphrase)

Pupil: Okay. Cause it determines the amount of focus.

Tutor: Right. (confirmation)

Pupil: Okay. And then the vitreous humor is just like the jelly – the sensitive or light that goes like straight through.

Tutor: Correct. (confirmation)

Pupil: And then the fovea is- The cones are in the fovea.

Tutor: Right. (confirmation)

Pupil: Cause the amount of light or the sensitivity and the blind spot – like the peripheries and the sclera is like just an outside covering.

Tutor: Right. (confirmation)

4.2.4 Concept elaboration

In concept elaboration, participants clarified text terminology and explained concepts. These episodes contained elaborations which were developed and explained over multiple statements. However, there was little sense in these episodes that the speaker was trying to develop their knowledge or resolve confusion. There were few self-monitoring statements in these episode indicating doubt or confusion. Instead, they seemed to be simply sharing information that they already knew or had figured out during the initial reading phase. For these reasons, concept elaboration episodes might be considered a hybrid of knowledge-telling and knowledge-building.

Excerpt D provides an example of a high-performing general tutor who was explaining to the tutee about the indirect pathway of neural signals through the retina. One thing to notice about this example is that the tutor does not display any doubt or confusion about his or her explanations. The tutor seems to deliver the explanation quite confidently.

[D] Tutor: Then there are the amacrine cells, which connects bipolars to bipolars or ganglions to ganglions. (paraphrase)

Pupil: Is that the indirect path?

Tutor: Yeah. An indirect path would be through a receptor. Through a horizontal to another receptor (elaboration)

Tutor: To a bipolar. To another bipolar, through an amacrine. (elaboration)

Tutor: And then, back to the bipolar again. (elaboration)

Tutor: And then to the ganglion. Through amacrine to another ganglion. (elaboration)

Tutor: And then out. (elaboration)

Pupil: Oh my gosh. I'll have to remember that.

Tutor: Like anything that's not the receptor, bipolar, ganglion is an indirect path. (elaboration)

Tutor: Once you get the horizontal which is receptor to receptor (paraphrase)

Tutor: or amacrine which is bipolar to bipolar, or ganglion to ganglion, (paraphrase)

Tutor: you have an indirect path. (elaboration)

Pupil: Oh, good. Okay. That makes sense.

4.2.5 Elaborated review

When reviewing, some participants made additional new connections with prior knowledge, and generated novel examples and analogies to further develop the meaning of concepts. Thus, instead of simply repeating the same ideas again, these episodes involved rethinking or application of the ideas. Such elaborations may have reflected new ideas that the tutor generated while reviewing, which had not occurred to them the first time they had discussed the topic. Thus, these episodes were considered to be clear instances of knowledge-building.

Excerpt E provides an example of elaborated reviewing in which a high-performing explanation tutor reviewed the cornea in response to a tutee question and assertions. In this episode, the tutor elaborated by making inferences about the shape of the cornea, drawing connections to the lens, and making an analogy to a camera.

[E] Pupil: So the cornea directs the light onto the lens. The cornea is 70 percent of the focusing?

Tutor: Yeah.

Pupil: And then the ciliary muscles stretch or contract the lens, focuses it onto the main — the most part of the light onto that fovea, which contains a lot of the cones.

Tutor: Yeah, yeah, that's pretty much how it works. (paraphrase)

Tutor: And since the cornea is—The shape of the cornea doesn't change. (elaboration)

Pupil: Okay.

Tutor: It's always the same shape. (elaboration)

Tutor: So even though that'd account for most of the focusing power, (paraphrase)

Tutor: [the lens] accounts for any variables, (elaboration)

Tutor: like movement and distance. (elaboration)

Tutor: Since this is attached to ciliary muscles (paraphrase)

Pupil: Okay.

Tutor: Kind of like the focus on a camera. (elaboration)

4.2.6 Sense-making

In sense-making episodes, participants drew upon prior knowledge and generated inferences in order to address perceived errors and confusions (or perhaps their own curiosity). Such episodes were characterized by frequent uncertainty and guesswork, and learners were often trying to use available information to speculate about plausible explanations. As a result, the conclusions reached in sense-making episodes were not always correct, but the mental effort put into the sense-making process could be beneficial. These episodes were very similar in nature to self-explaining (Chi, 2000) and provide clear instances of reflective knowledge-building.

Excerpt F shows a high-performing question tutor who was trying to understand how muscles in the eye (e.g. iris muscles and ciliary muscles) are regulated in order respond to changing amounts of light and other inputs. He drew upon his prior knowledge of muscles and nerves to infer that there must be some additional neural pathway from the brain to the eye that sends these regulatory signals.

[F] Tutor: Yeah. But it makes sense that- (monitoring)

Tutor: that once it gets to the brain that it probably would have some type of control system (elaboration, monitoring)

Tutor: where the brain's like okay you're giving me this. (elaboration)

Tutor: We need something to change. (elaboration)

Pupil: Something to change.

Tutor: So yeah, then it would probably—(monitoring)

Tutor: All of these are probably connected by some other nerves going back and controlling (elaboration, monitoring)

Tutor: that cause all- every muscle needs to be controlled by some nerve (elaboration)

Pupil: Yeah. So it needs to have something to say- to determine whether its- like what's going on there.

Tutor: Exactly.

Table 11. Raw mean episode frequencies for high and low performers by condition

	Conditions			
	High Performers			
Episode Category	Active Reviewers	General tutors	Explain tutors	Question tutors
Summary	11.3 (2.2)	11.7 (5.7)	10.8 (2.5)	10.7 (3.3)
Review	9.8 (7.5)	5.0 (1.8)	5.8 (2.8)	5.3 (2.9)
Passive feedback		2.0 (1.5)	0.8 (1.0)	2.7 (1.6)
Concept elaboration	1.5 (1.6)	5.0 (5.4)	5.3 (2.5)	3.5 (2.2)
Elaborated review	1.0 (1.5)	1.7 (1.2)	1.7 (1.4)	1.5 (0.8)
Sense-making	1.2 (1.8)	0.7 (1.2)	2.3 (2.5)	1.5 (1.8)
	Low Performers			
	Active Reviewers	General tutors	Explain tutors	Question tutors
Summary	11.7 (3.9)	14.0 (3.4)	14.7 (3.4)	13.2 (4.4)
Review	6.7 (5.0)	5.2 (2.8)	5.2 (4.8)	8.0 (4.3)
Passive feedback		2.5 (1.9)	5.2 (2.1)	2.7 (1.6)
Concept elaboration	0.3 (0.5)	1.7 (1.2)	1.8 (1.5)	3.0 (2.6)
Elaborated review	0.0 (0.0)	0.3 (0.5)	0.0 (0.0)	0.0 (0.0)
Sense-making	0.5 (0.8)	0.3 (0.5)	0.5 (0.8)	0.5 (0.5)

4.3 LEARNING BEHAVIORS IN THE ACTIVE REVIEWING GROUP

Both high and low-performing reviewers displayed a knowledge-telling bias. Summary, review, and concept elaboration episodes occurred most frequently, and accounted for an average of 92.7% (*SD* = 11.2%) of reviewers' episodes. Elaborated review and sense-making activities accounted for only 7.3%. This difference was significant, $t(11) = 13.15$, $p < .001$. Because these two knowledge-building activities occurred so rarely, they were collapsed into a single category called "knowledge-building" for greater power in subsequent analyses.

To evaluate the influence of knowledge acquired during the studying phase (i.e. prior knowledge), correlations were computed between mid-test Questions Test scores and episode proportions. The four comprehension questions on this test required both recall and integration to answer, and so this measure was a good indicator of post-studying fact knowledge and understanding. Prior knowledge was negatively related to the proportion of summarizing, $r(10) = -.71$, $p = .01$; and reviewing, $r(10) = -.22$, *ns*.; but was positively related to concept elaboration, $r(10) = .79$, $p = .002$; and knowledge-building, $r(10) = .74$, $p = .006$. Thus, reviewers with higher prior knowledge spent more time engaged in elaborative activities during the think-aloud.

Differences in how high and low-performing active reviewers studied the material during the reviewing phase were then assessed in a one-way ANCOVA. Due to the strong relationships with prior knowledge, mid-test Questions Test scores were included as a covariate. Mean proportions of episodes are given in Table 12. There were no significant differences between high and low performers across episodes. However, there were significant effects of prior knowledge for concept elaboration, $F(1,9) = 13.38$, $p = .005$, $R^2 = .60$; and knowledge-building $F(1,9) = 14.64$, $p = .004$, $R^2 = .62$.

Table 12. Adjusted mean proportions of episodes for active reviewers

Episode Category	High Reviewers ($n = 6$)	Low Reviewers ($n = 6$)
Summary	.52 (.12)	.58 (.12)
Review	.43 (.19)	.24 (.19)
Concept elaboration	.02 (.04)	.05 (.04)
Knowledge-building	.02 (.09)	.12 (.09)

An interpretation of these results is that active reviewers' prior knowledge or reading skill may have constrained their knowledge-building activities. Reviewers who had learned more during the first phase had more information to draw upon, and thus were better equipped to further build on this knowledge. In fact, it seems plausible that the most successful reviewers were spontaneous self-explainers. These students likely engaged in knowledge-building during the studying phase, leading to a stronger initial understanding. During the think-aloud phase, they reviewed what they had previously learned to reinforce their understanding, and likely acquired or generated additional knowledge along the way. In contrast, reviewers with low prior knowledge were probably using shallow strategies throughout the study. As a result, they learned less in the initial studying phase, and then lacked a necessary foundation for knowledge-building. During the think-aloud session, they continued to struggle with basic concepts and relied on shallow strategies. When they did attempt to engage in knowledge-building, they may have been less able to generate meaningful elaborations and deductions.

4.4 LEARNING BEHAVIORS IN THE PEER TUTORING GROUPS

Peer tutors also displayed a strong knowledge-telling bias. 92.7% (SD = 7.6) of tutors' behaviors on average were knowledge-telling episodes, and only 7.3% were knowledge-building episodes, $t(35) = 33.47$, $p < .001$. There were no differences across tutoring conditions. Due to the low frequency of elaborated reviewing and sense-making episodes, these behaviors were again collapsed into a single "knowledge-building" category.

The influence of prior knowledge was tested by calculating correlations between mid-test Questions Test scores and episode proportions. Prior knowledge was negatively related to the proportion of summary, $r(34) = -.46$, $p = .005$; and passive feedback episodes, $r(34) = -.32$, $p = .056$. There was no relationship to review episodes, $r = .00$. Prior knowledge was positively related to concept elaboration, $r(34) = .54$, $p = .001$, and knowledge-building, $r(34) = .45$, $p = .006$. Thus, peer tutors who entered the tutoring session with stronger prior knowledge were more likely to engage in knowledge-building and less likely to engage in knowledge-telling.

Differences in high and low-performing tutors' behaviors were compared in a 2 (performance) x 3 (condition) ANCOVA, taking into account condition and controlling for prior knowledge. There was no main effect of experimental condition on any of the variables, although there was a performance by condition interaction for passive feedback, $F(1,29) = 5.18$, $p = .012$, $R^2 = .26$. High-performing explanation tutors produced a low proportion of these episodes whereas low-performers produced a higher proportion. Due the overall lack of differences among the conditions, Table 13 reports only collapsed data for high and low-performing tutors.

Overall, high performers engaged in more knowledge-building than low-performing tutors, $F(1,29) = 14.67$, $p = .001$, $R^2 = .34$, with no significant effect of prior knowledge. These results suggest that knowledge-building activities contributed to tutor learning during the tutoring phase. Low performers engaged in more passive feedback than high performers, $F(1,29) = 4.70$, $p = .039$, $R^2 = .14$; again with no significant effect of prior knowledge. When tutors stopped making substantive contributions to the discussion, instead focusing on tutees' knowledge and understanding, tutor learning appeared to be hindered. An implication of this finding is some tutors may be unable to handle the complexity of monitoring and regulating both their own knowledge and their knowledge of their tutee.

High and low performers did not differ in the proportion of concept elaboration, but these episodes were influenced by prior knowledge, $F(1,29) = 4.98$, $p = .034$, $R^2 = .15$. Peer tutors who knew more coming into the tutoring sessions engaged in more concept elaboration. This finding makes sense if tutors were not generating many novel elaborations in these episodes. Instead, tutors were using the knowledge they had gained during initial studying phase. Thus, as captured in the episode coding scheme, concept elaboration episodes most likely represented artifacts of earlier knowledge-building rather than knowledge-building during tutoring. Thus, they probably did not contribute as much to learning-by-tutoring.

Table 13. Adjusted mean proportions of episodes for tutors

Episode Category	High Tutors (n = 18)	Low Tutors (n = 18)	Cohen's d effect size
Summary	.47 (.17)	.50 (.17)	-0.18
Review	.20 (.13)	.23 (.13)	-0.23
Passive feedback	.07 (.08)*	.13 (.08)	-0.75
Concept elaboration	.14 (.12)	.12 (.12)	0.20
Knowledge-building	.12 (.07)*	.02 (.07)	1.43

These analyses suggested that tutors' knowledge-building activities played a role in tutor learning above and beyond the influence of prior knowledge. To further probe this idea, regression analyses were conducted to more directly look at prior knowledge and knowledge-building as predictors of scores on specific post-test measures. Table 14 summarizes the results of these analyses. Not surprisingly, prior knowledge was a strong predictor of post-test scores. However, peer tutors knowledge-building activities made an additional and meaningful contribution to Questions Test scores, which assessed deeper understanding of the material.

Table 14. Regression coefficients predicting post-test scores

Measure	Prior knowledge				Knowledge-building			
	B (SE)	*β*	*t*	*p*	***B*** (SE)	*β*	*t*	*p*
Definition Test	.58 (.24)	.41	2.42	.021	14.5 (20.7)	.12	.70	*ns*
Questions Test								
4 core questions	.69 (.12)	.69	5.97	< .001	18.4 (9.9)	.22	1.86	.071
6 new questions	.62 (.11)	.66	5.52	< .001	18.6 (9.7)	.23	1.92	.064
Total questions	1.3 (.19)	.72	6.99	< .001	37.9 (16.1)	.24	2.36	.024

4.5 SUMMARY

In support of Hypothesis A, knowledge-building activities were associated with stronger learning outcomes for both peer tutors and students who reviewed the text on their own. However, both reviewers and tutors, and high and low performers, displayed a strong knowledge-telling bias (Hypothesis B). Participants showed a much higher mean frequency and proportion of knowledge-telling episodes. These activities were not associated with effective learning, and in some cases seemed to actually hinder learning, as predicted by Hypothesis C.

There were almost no differences among tutors in the three tutoring conditions. This means that introducing tutors to strategies that were known to be effective, either framed as explanation or question-answering strategies, had little effect without extensive training (Hypothesis D). These data argue against the idea that peer tutors focus on knowledge-telling strategies simply because they are not aware of knowledge-building strategies. Rather than lack of awareness, lack of ability or willingness may be more important factors.

Interestingly, there appeared to be meaningful differences in the role of prior knowledge in learning from tutoring versus reviewing a text. Reviewers' knowledge-building activities were dependent upon their prior knowledge. Active reviewers who learned more early on were able to continue to build upon this knowledge and further advance their understanding. Reviewers who learned less during the initial study phase were more constrained in their ability to engage in knowledge-building. It is likely that high performing reviewers may have been high-skilled readers to begin with, and perhaps spontaneously self-explained throughout the study.

Only one tutor behavior, concept elaboration, was primarily dependent to prior knowledge. In these episodes, tutors shared elaborations that were probably generated during the reading phase. These early insights helped them do better on the mid-test questions, and that knowledge carried over to the post-test. However, several tutoring behaviors differed across high and low performing tutors and were not as dependent on prior knowledge. Most importantly, knowledge-building activities (i.e. elaborated reviewing and sense-making) seemed to be accessible and beneficial to tutors with both more and less pre-tutoring knowledge.

An important implication of these results is that the mechanisms of learning may differ in peer tutoring versus solo reading. Students who studied a text alone seemed limited by their own personal resources and skills. Prior knowledge was also critical for peer tutors; tutors with more knowledge could use this knowledge to give more elaborated explanations and reason about the subject matter. However, it appeared that the peer tutoring process also provided tutors with learning opportunities that rose above such constraints. This result could help to explain why tutor learning has been observed across diverse students, such as tutors with learning disabilities (Mathes & Fuchs, 1994).

One way that this might have occurred is via tutor-tutee interactions. Whereas reviewers are working alone, tutors must work with and accommodate the needs of their tutee. These tutees might introduce new ideas or explanations that push the tutor to further develop their explanations (Roscoe & Chi, in-press-A). In particular, tutees may have asked questions that prompted the tutor to rethink and elaborate the material, thus providing a form of unintentional scaffolding for tutors' knowledge-building. Even when tutors understood the material, tutee questions could have pushed the tutor to find novel ways to explain concepts. Similarly, tutee questions might have provided cues that tutors' explanations were incomplete or incorrect, leading the tutor to revise their knowledge. This hypothesis is considered in the next section.

5.0 ANALYSES AND RESULTS: RESPONDING TO TUTEE QUESTIONS

Prior research has found that tutee questions had a meaningful impact on tutors' subsequent answers and explanations (Roscoe & Chi, in-press-A). Deep questions elicited deeper and more metacognitive tutor responses, whereas shallow responses elicited shallow and non-metacognitive responses. One unanswered question in that prior work was whether different question-response patterns were directly related to tutor learning.

5.1 CODING OF QUESTION-RESPONSE EXCHANGES

The episodes identified in prior analyses could contain multiple tutee questions, and so the tutoring protocols were resegmented using "question-response exchanges" as the unit of analysis. Tutee questions were defined as interrogative statements in which tutees requested information or verification of information. Questions about task procedures or off-topic issues were excluded. Tutee questions were then coded as "shallow" or "deep." Deep questions 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.

Tutor responses were defined as any information or feedback provided by the tutor in answer to the question. Tutor responses were coded as shallow or deep depending on whether they contained elaboration statements. Tutor responses were further coded as "metacognitive" or “non-metacognitive” based on whether they contained overt self-monitoring statements (e.g. “I don’t know” or “I’m guessing”). Due to the finding that tutors in the different tutoring groups did not differ in their learning outcomes or episode-level behaviors, data from the tutoring conditions were collapsed for these analyses. Mean frequencies are given in Table 13.

A total of 772 questions were asked across all tutees. Reliability was assessed by having two raters (author and a post-doctoral researcher) code the same set of ten complete protocols. There was 84.2% agreement for coding of question depth, 96.3% agreement for coding of response depth, and 96.8% agreement for coding of response monitoring. Overall, 72.7% of questions were shallow and 27.3% were deep, which is comparable to results from in a prior study (Roscoe & Chi, in-press-A).

Excerpts G and H provide examples of two question-response exchanges. Excerpt G gives an example of a deep/deep exchange (i.e. deep question/deep response) that is also a deep/meta exchange (i.e. deep question/metacognitive response). Excerpt H gives an example of a shallow/shallow exchange that is also a shallow/non-meta exchange.

[G] Pupil: I know we have of course a lot of cones in the fovea and the number attenuates as you go out. Why don’t we just have exclusively cones? (deep question)

Tutor: Because then we couldn’t see at low light levels. Because cones need or require a lot of light to be stimulated. Yeah, rods are, like, rods are essentially almost black and white vision, but they can go really, really good night - that’s what gives you night vision is the rods. Your cones are worthless at night. Yeah, that’s the real reason. As far as I understand. (deep response, metacognitive)

[H] Pupil: Oh. What’s the retina do? (shallow question)

Tutor: The retina contains all those photoreceptors. (shallow response, non-metacognitive)

5.2 QUESTIONS, RESPONSES, AND TUTOR LEARNING

The mean percentages of tutor responses by question type were examined, collapsing across high and low performers. On average, 54.3% (*SD* = 36.4%) of deep tutee questions elicited a deep tutor response, whereas only 13.3% (*SD* = 19.4%) of shallow questions received a deep response, $t(35) = 7.62$, $p < .001$. Similarly, 44.2% (*SD* = 31.2%) of deep questions elicited a metacognitive response, whereas only 13.8% (*SD* = 10.1%) of shallow questions received a metacognitive response, $t(35) = 5.27$, $p < .001$. These results replicated patterns reported by Roscoe & Chi (in-press-A). As predicted by Hypothesis E, deep questions helped to elicit knowledge-building responses, whereas shallow questions elicited shallow responses.

A second analyses considered whether the frequency of shallow or deep questions were linked to stronger learning outcomes. In order to evaluate the possible influence of prior knowledge, correlations were computed between response patterns and Questions Test mid-test scores. There were no significant correlations with prior knowledge, although deep questions were somewhat positively associated with mid-test scores, $r(58) = .29$, *ns*.; and shallow questions were negatively associated with these scores, $r(58) = -.32$, $p = .059$. Tutors with stronger prior knowledge were somewhat more likely to be asked a deeper question by tutees. This may be because tutors with more prior knowledge offered more concept elaboration explanations, which then prompted better tutee questions.

Subsequently, high and low performers were compared in a one-way ANCOVA, controlling for prior knowledge. The frequency of deep and shallow questions asked did not differ for high and low-performers (Table 15). There were no effects for prior knowledge. Thus, the depth of questions did not seem to directly influence tutor learning.

Table 15. Raw mean frequencies of questions and question-response exchanges

Question and response types	High tutors (n = 18)	Low tutors (n = 18)
Question type		
Deep questions	6.8 (4.2)	5.0 (4.6)
Shallow questions	11.7 (7.3)	19.4 (10.8)
Response depth		
Deep/deep exchanges	5.1 (3.3)	1.6 (2.0)
Deep/shallow exchanges	1.7 (2.0)	3.3 (3.1)
Shallow/deep exchanges	2.1 (2.1)	0.8 (0.9)
Shallow/shallow exchanges	9.7 (6.6)	18.6 (10.5)
Response metacognition		
Deep/meta exchange	2.9 (2.7)	2.3 (2.5)
Deep/non-meta exchanges	3.8 (2.7)	2.7 (2.5)
Shallow/meta exchanges	1.8 (1.5)	2.4 (1.9)
Shallow/non-meta exchanges	10.1 (6.7)	17.0 (10.0)

A more important issue is how tutors answered these questions. Each question could be either deep or shallow, and subsequent responses could also be deep or shallow. To examine tutors' response patterns, I examined the proportion of deep responses that tutors gave in answer to deep questions, and the proportion of deep responses to shallow questions. Such proportions assessed the extent to which tutors gave elaborative responses to different question types. Giving deep responses to questions should be beneficial to tutor learning. Because question-responses were coded dichotomously (i.e. deep vs. shallow), only data for deep responses are reported.

First, the influence of prior knowledge on tutors' responses was tested by computing correlations between proportions of different question-response patterns and mid-test Questions Test scores. Prior knowledge was significantly and positively related to the proportion of deep responses given to deep questions (i.e. deep/deep exchanges), $r(34) = .45$, $p = .006$; and the proportion of deep questions given in answer to shallow questions (i.e. shallow/deep exchange), $r(34) = .46$, $p = .005$. Thus, tutors with more prior knowledge were more likely or able to give deeper responses to various questions.

Differences in high and low-performing tutors' response pattern proportions were then examined in a one-way ANCOVA, controlling for prior knowledge (Table 16). High performers exhibited a higher proportion of deep/deep exchanges than low performers, $F(1,33) = 16.15$, $p < .001$, $R^2 = .33$, with no significant effect of prior knowledge. The proportion of shallow/deep exchanges did not differ significantly across high and low performers, but was somewhat influenced by prior knowledge, $F(1,33) = 3.34$, $p = .077$, $R^2 = .09$. Thus, peer tutors with stronger prior knowledge may have offered more elaborated responses to shallow questions. These shallow/deep exchanges probably did not reflect knowledge-building, but rather the sharing of ideas learned during the studying phase, similar to concept elaboration.

Table 16. Adjusted mean proportions of question-response exchanges

Question-response exchanges	High tutors (n = 18)	Low tutors (n = 18)	Cohen's d effect size
Response depth			
Deep/deep	.78 (.32)*	.30 (.32)	1.50
Shallow/deep	.16 (.20)	.10 (.20)	0.30
Response metacognition			
Deep/meta	.46 (.37)	.42 (.37)	0.11
Shallow/meta	.14 (.12)	.13 (.12)	0.08

In sum, peer tutors certainly found it easier to give deeper responses if they had more prior knowledge to draw from. However, in support of Hypothesis F, learning was still enhanced for both high and low knowledge tutors if they tried to answer deep questions with deep responses. In other words, this form of tutor-tutee interaction did indeed provide tutor learning opportunities that were somewhat dissociable from tutors' prior knowledge and skills. This was further explored in regression analyses in which prior knowledge and deep/deep exchanges were evaluated as predictors of post-test scores. Prior knowledge was entered first, followed by deep/deep exchanges. These analyses showed deep/deep exchanges were significant predictors of tutors' final understanding, above and beyond prior knowledge (Table 17).

Table 17. Regression coefficients predicting post-test scores

	Prior knowledge				Deep/deep exchanges			
Measure	***B*** (SE)	*β*	*t*	*p*	***B*** (SE)	*β*	*t*	*p*
Definition Test	.45 (.23)	.32	1.96	.059	.08 (.04)	.34	2.08	.046
Questions Test								
4 core questions	.68 (.11)	.67	5.96	<.001	.04 (.02)	.24	2.2	.038
6 new questions	.58 (.11)	.61	5.48	<.001	.06 (.02)	.33	2.96	.006
Total questions	1.26 (.18)	.68	7.11	<.001	.10 (.03)	.31	3.27	.003

Similar analyses were also conducted to examine differences in metacognitive and non-metacognitive tutor responses. There were no significant differences in the proportion of deep/meta or shallow/meta exchanges, nor were there effects of prior knowledge. These results suggest that tutors' self-monitoring in response to questions was not strongly related to learning outcomes. The depth of the questions and responses seemed to matter more.

One possible explanation of this result is that tutors' self-evaluations did not focus on comprehension, but instead attended to more trivial matters such as their ability to remember a word. Similarly, high and low performing tutors may have responded differently to these self-judgments. Some tutors may have monitored themselves accurately and recognized crucial knowledge gaps, but then failed to follow up with any kind of knowledge elaboration or revision. The following analyses, in the next sections, dig deeper into peer tutors' metacognitive activities and their relationship to learning outcomes.

6.0 ANALYSES AND RESULTS: METACOGNITIVE MONITORING

In the theoretical framework which has been outlined, processes of self-monitoring should be an important component of knowledge-building and tutor learning. As tutors give explanations and respond to questions, they may monitor the correctness, coherence, and relevance of their statements. They can also reflect on whether their explanations or the source materials make sense. Such self-evaluations provide important information to the learner about the state of their knowledge. When tutors realize that they have erred or are confused, they might use knowledge-building strategies to address these problems. In doing so, their understanding would be improved. The following analyses further explore these hypotheses.

6.1 CODING OF METACOGNITIVE BEHAVIORS

Roscoe & Chi (in-press-A) analyzed metacognitive monitoring at the level of episodes of verbal explanations (defined in the same manner as episodes in the current study). Episodes were coded as metacognitive or non-metacognitive depending on whether the episode contained overt monitoring statements. Such statements consisted of short phrases in which students evaluated some aspect of their cognition, such as comprehension or accuracy. Although participants probably engaged in covert metacognitive monitoring, the occurrence of overt statements provided the only direct evidence that monitoring had occurred. In the current project, the first step was to replicate this previous episode-level method of analysis.

Subsequent analyses probed tutors' metacognitive behaviors more deeply by examining the content, valence, and context of individual metacognitive statements. Raw means for all analyses are provided in Table 18. Data from all three tutoring conditions were again collapsed.

Table 18. Raw mean frequencies of metacognitive behaviors

Metacognitive activities	High tutors (*n* = 18)	Low tutors (*n* = 18)
Episodes		
Metacognitive episodes	13.6 (4.0)	12.1 (3.8)
Non-metacognitive episodes	12.6 (5.2)	14.2 (4.8)
Content of statements		
Comprehension-monitoring	17.7 (8.3)	14.1 (8.1)
Memory-monitoring	4.3 (2.4)	4.3 (2.6)
Accuracy-monitoring	3.9 (2.3)	2.3 (1.6)
Attention-monitoring	1.4 (1.5)	1.6 (1.8)
Valence of statements		
Positive valence	7.9 (3.8)	6.4 (3.3)
Uncertainty valence	9.4 (4.9)	6.6 (3.8)
Negative valence	10.2 (4.1)	9.3 (5.1)
Context of statements		
Elaborative context	14.2 (8.1)	5.2 (4.8)
Non-elaborative context	13.4 (6.7)	17.1 (7.8)

The content of a statement referred to the aspect of cognition being evaluated. Four content categories emerged: comprehension, memory, accuracy, and attention. Comprehension-monitoring statements focused on tutors' understanding of the material ("I got that" or "I don't know") or their ability to understand the material ("This is easy" or "This is where it gets complicated"). Memory-monitoring statements related to tutors' recall ("I forget" or "I don't remember") or their ability to remember ("I'll never remember this for the test" or "That's easy to remember"). Accuracy statements evaluated correctness ("I got that right" or "I think I told you the wrong thing"). Finally, attention statements indicated what tutors felt was or was not worth attending to because of importance ("That's all you need to know") or interest ("This is interesting"). Thus, this analysis was informative the types of judgments tutors were making as they taught. Reliability was assessed by having two coders (the first author and a graduate student) code the same set of ten complete protocols. The second rater was given a sheet that described the categories and provided examples, which was accompanied by a brief verbal explanation of the coding. Agreement between the raters was adequate (85.8%).

Metacognitive statements could also be assigned a valence, which referred to whether the statement was a positive, negative, or uncertain evaluation. Positive statements indicated that the tutor felt they understood the material and their ideas were correct, whereas negative statements indicated perceived comprehension failures or mistakes. Uncertainty statements ("I'm guessing" or "I'm not sure, but I think") occurred when the tutor felt they only partially understood the material. Overall, this analysis indicated when tutors felt they were or were not learning. The valence of statements was largely unambiguous and reliability data was not collected.

A final analysis assessed whether statements were associated with elaboration. For example, in one case tutors might generate an elaboration and then monitor the quality of their understanding. In another example, tutors might express uncertainty, and then offer an elaboration to address the confusion. Thus, this coding enables an assessment of how and whether tutors were actually using metacognitive monitoring to guide their knowledge-building activities. If the monitor statement occurred in conjunction with elaboration, it was coded as "elaborative context." If the statement did not occur in conjunction with elaboration statements, it was coded as "non-elaborative context."

6.2 METACOGNITIVE MONITORING AND TUTOR LEARNING

The proportion of metacognitive episodes was only weakly and positively correlated with prior knowledge, $r(34) = .14$, *ns*. The proportion of metacognitive episodes also did not differ across high and low performing tutors, and there was not an effect for prior knowledge. Thus, high and low-performing tutors seemed to monitor themselves at about the same level. Roughly half of their episodes contained overt monitoring statements on average (Table 19).

Peer tutors' monitoring activities were next examined at the more fine-grained statement level (Table 19). In terms of content, comprehension-monitoring statements occurred most frequently (62%) and other types of monitoring occurred less often. None of the content categories were significantly correlated with prior knowledge. High or low-performing tutors' also did not differ in their proportion of comprehension, memory, accuracy, or attention monitoring, and their were no effects of prior knowledge. These results indicated that peer tutors were attending to matters of understanding rather than only testing their memory and accuracy. However, the lack of differences between high and low performers suggest that tutor learning outcomes could not be directly explained by a failure to monitor one's own learning.

Table 19. Adjusted mean proportions of metacognitive activities

Metacognitive activities	High tutors (n = 18)	Low tutors (n = 18)	Cohen's d effect size	Correlation w/ proportion of KB
Episodes				
Metacognitive episodes	.53 (.18)	.46 (.18)	0.39	.37*
Content of statements				
Comprehension-monitoring	.62 (.20)	.62 (.20)	0.00	.34*
Memory-monitoring	.18 (.13)	.19 (.13)	-0.08	-.31
Accuracy-monitoring	.14 (.10)	.13 (.10)	0.10	-.23
Attention-monitoring	.05 (.08)	.07 (.08)	-0.25	-.09
Valence of statements				
Positive valence	.28 (.14)	.31 (.14)	-0.21	-.24
Uncertainty valence	.35 (.14)	.27 (.14)	0.57	.29
Negative valence	.37 (.11)	.42 (.11)	-0.45	-.08
Context of statements				
Elaborative context	.47 (.21)*	.25 (.21)	1.05	.64*

In terms of valence, none of the statement types (i.e. positive, uncertain, or negative) were significantly correlated with prior knowledge. There were also no differences across high and low performers in the frequencies of these statements, and no effects of prior knowledge. Thus, high and low performers did not seem to differ greatly in their perceptions of understanding. It is worth noting that negative and uncertain valence statements were the most common (approximately 70%). These results suggest that peer tutors were reasonably aware of their mistakes and confusions, and high and low performers were about even in error detection.

Taken together, results at the episode and statement levels showed that high and low performing tutors monitored themselves in similar ways. These findings were surprising in light of prior research and theoretical arguments for the role of metacognitive monitoring in tutor learning. It was hypothesized that high performing tutors would have monitored themselves more often (i.e. more monitoring statements) and/or appropriately (i.e. greater attention to comprehension and comprehension failures) than low performers. This was not the case, which seemed to provide evidence counter to the metacognitive hypothesis.

Another important issue to consider, however, is the relationship between peer tutors' self-monitoring and their knowledge-building behaviors. Although monitoring may not directly contribute to learning of the material, the detection of knowledge deficits should still provide opportunities for knowledge-building to occur. In fact, positive and significant relationships were observed between several monitoring activities and knowledge building. Positive partial correlations (controlling for prior knowledge) were found between the proportion of knowledge-building episodes and metacognitive episodes; $r(33) = .37$, $p = .027$, and comprehension-monitoring statements, $r(33) = .34$, $p = .049$. Knowledge-building also seemed to be negatively related to memory-monitoring, $r(33) = -.31$, $p = .068$. These results provide partial support for Hypotheses G and H. Overall monitoring and comprehension-monitoring were related to knowledge-building, although they were not directly related to learning outcomes.

One particular aspect of knowledge-building, sense-making, may have been especially sensitive to metacognitive monitoring. This behavior involved making sense of confusing ideas or filling perceived knowledge gaps. Thus, the knowledge-building category was broken down into elaborated reviewing and sense-making categories, and additional partial correlations were computed. Overall, the proportion of elaborated reviewing was not significantly correlated with metacognitive activities, although may have been negatively correlated with accuracy monitoring, $r(33) = -.31$, $p = .073$. As expected, the proportion of sense-making was significantly and positively correlated with metacognitive episodes, $r(33) = .37$, $p = .027$; comprehension-monitoring, $r(33) = .37$, $p = .027$; and uncertain valence statements, $r(33) = .36$, $p = .034$. Sense-making was negatively related to memory monitoring, $r(33) = -.38$, $p = .024$.

Overall, these results show that monitoring was in fact a meaningful component of knowledge-building. The detection of confusions or other barriers to understanding prompted some tutors to strive to elaborate or revise their knowledge. Thus, self-monitoring was a first step for some knowledge-building activities, and so a lack of self-monitoring could lead to a knowledge-telling bias. If tutors do not (or cannot) periodically evaluate their understanding, they may overlook important knowledge gaps and thus never realize that some deeper reorganization or rethinking of the material is warranted.

How can one reconcile the finding that high and low performers did not differ in self-monitoring, but yet differed in their knowledge-building activities? One possibility is that high and low performing tutors did not make use of their metacognitive evaluations in the same ways. Perhaps high performing tutors tried to address perceived confusions through knowledge-building elaborations, whereas low performing tutors responded to comprehension failures with more paraphrasing and rehearsal. In other words, differences between high and low performing tutors might have reflected a problem of self-*regulation* rather than self-monitoring.

The analysis of the context of tutors' monitoring statements directly addressed that idea. The proportion of monitoring statements occurring in an elaborative context was positively related to prior knowledge, $r(34) = .50$, $p = .002$. Thus, tutors with stronger knowledge were more likely to combine their metacognitive and knowledge-building activities. However, a one-way ANCOVA (controlling for prior knowledge) also showed that high performers exhibited a higher proportion of elaborative context statements than low performers, $F(1,33) = 7.55$, $p = .01$,

$R^2 = .19$. There was not a significant effect of prior knowledge. As hypothesized (Hypothesis J), peer tutors learned more when they generated and then evaluated meaningful elaborations, or responded to their self-monitoring by integrating information and generating inferences.

A regression analysis was used to further explore how well elaborative context monitoring predicted post-test scores (Table 20). The proportion of elaborative context statements contributed to post-test scores to some extent beyond the effects of prior knowledge. In particular, these statements were significantly related to scores on the 6 new questions on the post-test Questions Test, which tutors had never seen before. Thus, this self-regulatory activity may have helped tutors to build up a strong knowledge of the material that allowed them to think and reason with the text concepts in novel or more flexible ways.

Table 20. Regression coefficients predicting post-test scores

	Prior knowledge				Elaborative context statements			
Measure	***B*** (SE)	*β*	*t*	*p*	***B*** (SE)	*β*	*t*	*p*
Definition Test	.45 (.24)	.32	1.87	.070	11.7 (2.7)	.29	1.70	.098
Questions Test								
4 core questions	.73 (.12)	.72	5.86	<.001	3.4 (3.6)	.12	.95	.35
6 new questions	.58 (.11)	.61	5.11	<.001	8.3 (3.3)	.30	2.51	.017
Total questions	1.3 (.20)	.72	6.62	<.001	11.2 (5.8)	.21	1.94	.061

6.3 SUMMARY

Analyses of tutors' metacognitive behaviors extended and elaborated previous analyses of tutor learning behaviors. Both high and low-performing tutors monitored themselves in similar ways overall, and a lack of monitoring was not a particular barrier for tutors. Importantly, tutors focused most often on their comprehension of the material rather than more superficial self-evaluations, and were able to detect problems of comprehension. Such comprehension-monitoring may have facilitated knowledge-building. Thus, not only were tutors able to monitor themselves, they were able to attend to salient and important signs of their progress. As with other tutor behaviors, having stronger prior knowledge available made these processes easier for tutors. However, metacognitive monitoring was not completely constrained by this prior knowledge. One reason for this pattern is that the tutoring process naturally supported or prompted self-monitoring for most peer tutors. Prior analyses showed that deep tutee questions were one specific mechanism by which this could occur.

Most importantly, differences were observed in how peer tutors' used and responded to their self-evaluations. Tutors whose monitoring activity occurred in conjunction with elaboration learned more effectively, whereas tutors whose monitoring activity occurred without elaboration learned less effectively. These results are comparable to previous analysis showing that tutors who tried to give deep answers to deep questions scored higher on post-tests than tutors who only gave shallow answers. These patterns suggest that tutor learning depended on how tutors chose to respond to learning opportunities that arise. Deep tutee questions and self-evaluations of one's own understanding provided two such opportunities.

The question that emerges is why peer tutors chose to engage in knowledge-telling versus knowledge-building. Why did some tutors choose not to engage in elaborated reviewing, avoid give thoughtful answers to deep questions, or fail to address comprehension failures through sense-making? Prior knowledge and metacognitive abilities clearly influenced tutors' knowledge-building behaviors. These were not insurmountable obstacles, however, because the tutoring process itself naturally provided opportunities to overcome the limits of prior knowledge. One potential explanation is that peer tutors differed in their perceptions of their ability to engage in these tasks, or in their interest in doing so. Perhaps tutors who lacked self-efficacy might have avoided "making guesses" that could embarrass them. Similarly, tutors who found the material or task boring and useless might have been less willing to put forth extra effort to explore topics deeply. In other words, tutors may have avoided knowledge-building due to a lack of "motivation." These hypotheses are considered in the final set of analyses.

7.0 ANALYSES AND RESULTS: SELF-EFFICACY AND INTEREST

Self-efficacy and interest were assessed using two scales, the Attitudes Toward the Material Scale (ATMS) and the Attitudes Toward Tutoring Scale (ATTS). These measures produced four scores, which were analyzed via correlations with learning behaviors, and comparisons across high and low-performers. Of primary interest were the scores representing participants' motivation *immediately before* the tutoring sessions began. Thus, the current analyses focused on tutors' scores from the mid-test administration of the ATMS and ATTS.

7.1 CODING AND SCORING OF THE SCALES

Students responded to the scales by rating their agreement or disagreement with various self-efficacy and interest statements using a 7-point Likert-like scale. A "1" indicated stronger disagreement and a "7" indicated stronger agreement. Half of the items for each scale and subscale were positively worded and half were negatively worded. To score the items, tutors' ratings on the negative items were inverted. Thus, a "1" was changed to a "7," a "7" was changed to a "1" and so forth. Next, for each subscale the students' ratings for all 8 items were averaged to produce a single score. The range of possible values for these scores was between 1 and 7, with higher values indicating stronger self-efficacy or interest beliefs. Table 21 provides descriptive statistics for each of the scale scores.

Table 21. Descriptive statistics for motivation scale scores

Measure	Min	Max	Mean	SD
Learning efficacy	2.9	7.0	5.0	1.1
Topic interest	2.2	6.9	4.9	1.0
Tutoring efficacy	2.1	6.8	4.8	0.9
Tutoring interest	2.8	7.0	5.4	0.9

7.2 MOTIVATION AND LEARNING FROM TUTORING

Correlations between scale scores and mid-test Questions Test scores were first calculated to look at relationships with prior knowledge. Data from all 60 tutors were included. Prior knowledge was positively correlated with all measures, and significantly correlated with learning efficacy, $r(58) = .58$, $p < .001$; topic interest, $r(58) = .27$, $p = .037$; and tutoring efficacy, $r(58) = .32$, $p = .013$. Tutors who learned more from the initial studying phase felt more able to learn and the material, more able to teach it, and more interested in the topic (Table 22).

Partial correlations were then computed between the tutors' motivation scores and post-test assessments, controlling for prior knowledge (Table 20). Learning efficacy was positively correlated with scores on the 4 core questions, $r(57) = .37$, $p = .003$; and overall post-test Questions Test scores, $r(57) = .25$, $p = .058$. Similarly, topic interest was significantly and positively correlated with all post-test measures except for the 6 new questions. Topic interest was associated with Definitions Test scores, $r(57) = .29$, $p = .027$; the 4 core questions, $r(57) = .28$, $p = .03$; and overall Questions Test scores, $r(57) = .28$, $p = .03$.

Table 22. Correlations between motivation scale scores and post-test assessments

	Motivation assessment			
Knowledge assessment	Learning efficacy	Topic interest	Tutoring efficacy	Tutoring interest
Mid-test Questions Test	.59[a]	.27[c]	.32[c]	.14
Post-test Definitions Test	.04	.29[c]	.15	.20
Post-test Questions Test				
4 core questions	.38[b]	.28[c]	.28[c]	.15
6 new questions	.05	.20	.04	.24[d]
Total questions	.25[d]	.28[c]	.18	.22

Note. [a]$p < .001$. [b]$p < .01$. [c]$p < .05$. [d]$p < .07$. Correlations with mid-test Questions Test scores are zero-order correlations. Correlations with post-test assessments are partial correlations controlling for mid-test Questions Test scores.

Tutors' attitudes toward tutoring seemed to have less of an impact on learning outcomes. Teaching efficacy was only correlated with scores on the 4 core Question Test questions, $r(57) = .28$, $p = .034$. Correlations with other post-test measures were positive, but relatively low in magnitude. Tutoring interest was somewhat associated with scores on the 6 new questions, $r(57) = .24$, $p = .066$. All other correlations were positive and non-significant.

Several observations emerge from these patterns. One finding is that peer tutors' self-efficacy and interest beliefs seemed to have a meaningful relationship to their learning outcomes, which provides some support for Hypotheses K through N. Students who felt more capable of learning the material and found it enjoyable were better able to recall more information about eye components and answer comprehension questions testing integration and application of the material. These relationships were above and beyond the effects of prior knowledge. A second observation is that tutors attitudes' toward the material may have had a somewhat stronger impact on tutor learning than their attitudes towards tutoring. Learning efficacy and topic interest were more strongly correlated with learning outcomes than tutoring efficacy and tutoring interest. This may be important when designing programs with a goal of tutor learning. Special care might need to be taken to foster positive attitudes toward the topic during tutor training.

Motivational factors were linked most strongly to the 4 core questions that tutors had seen before on prior assessments. If students were engaged in the material and wanted to do well, it makes sense that they might focus their efforts on the concepts they knew would be on the final test. To put it another way, tutors who felt less capable and interested perhaps chose not try to further study these concepts, even though they knew they would be tested on it later.

7.3 MOTIVATION AND TUTOR LEARNING BEHAVIORS

Peer tutors' efficacy and interest beliefs were linked to learning outcomes. Of course, tutors did not learn simply by being "motivated." These motivational factors should exert their influence via tutors' behaviors. It was hypothesized that tutors' choices about how the cover the material, respond to questions, and respond to their own confusions should be linked to their self-efficacy and interest beliefs. These hypotheses were tested by considering relationships between motivation scores and peer tutors' learning behaviors in the tutoring sessions.

First, efficacy and interest scale scores were compared for the high and low-performing tutors to determine if there was an overall difference in motivation across performance level (Table 23). Although high performers showed higher mean scores than low-performers for all variables, these differences were not significant when controlling for prior knowledge. Prior knowledge was significantly related to learning efficacy, $F(1,33) = 4.49$, $p = .042$, $R^2 = .12$.

Table 23. Mean motivation scale scores for high and low-performing tutors

Motivation assessment	High tutors (n = 18)	Low tutors (n = 18)	Cohen's d effect size
Learning efficacy	5.4 (0.9)	5.0 (0.9)	0.44
Topic interest	5.2 (1.2)	4.6 (1.2)	0.50
Tutoring efficacy	5.3 (1.0)	4.6 (1.0)	0.70
Tutoring interest	5.7 (0.9)	5.2 (0.9)	0.55

Relationships between motivational factors and key tutor behaviors identified in previous analyses were then assessed via partial correlations, controlling for prior knowledge (Table 24). I included the activities which differed significantly across high and low performing tutors, and also showed some prediction of post-test scores above and beyond prior knowledge: proportion of knowledge-building episodes, proportion of deep/deep question-response exchanges, and proportion of monitoring statements in an elaborative context.

Table 24. Partial correlations between motivation assessments and key learning behaviors

	Motivation assessment			
Key activity	Learning efficacy	Topic interest	Tutoring efficacy	Tutoring interest
Knowledge-building episodes	.29	$.40^{c}$	$.44^{b}$	$.39^{c}$
Elaborated reviewing	.24	.01	$.35^{c}$	.18
Sense-making	.22	$.50^{b}$	$.33^{d}$	$.38^{c}$
Deep/deep question exchanges	.11	.18	.09	.13
Elaborative context statements	.06	$.34^{c}$	.12	$.39^{c}$

Note. $^{a}p < .001$. $^{b}p < .01$. $^{c}p < .05$. $^{d}p < .07$.

The proportion of knowledge-building episodes was significantly and positively correlated with topic interest, $r(33) = .40$, $p = .018$; tutoring efficacy, $r(33) = .44$, $p = .009$; and teaching interest, $r(33) = .39$, $p = .021$. Thus, peer tutors who felt more interested and capable engaged in significantly more knowledge-building than tutors with lower interest and efficacy.

To further explore these results, the two knowledge-building episode categories were reanalyzed separately. These activities could potentially differ in the motivational pressures that they place on tutors. In particular, sense-making could involve admitting one's mistakes or confusion and then trying to generate a plausible explanation through reasoning. The effortful nature of this task, and the potential for embarrassment (i.e. admitting ignorance and/or failing to come up with a good explanation) might have made motivational factors more salient for sense-making. Partial correlations showed that elaborated reviewing was correlated only with tutoring efficacy, $r(33) = .35$, $p = .041$. Tutors who felt more confident in their teaching abilities were more willing to generate new elaborations while reviewing. However, sense-making was positively correlated with topic interest, $r(33) = .50$, $p = .002$ and tutoring interest, $r(33) = .38$, $p = .025$; and also related to tutoring efficacy, $r(33) = .33$, $p = .053$. Thus, tutors who felt that the subject matter and tutoring task were enjoyable and valuable were more willing to confront their own confusions and try to generate new ideas to repair these deficits.

The proportion of deep responses to deep questions was only weakly related to any of the motivation assessment scale scores, and none of these relationships were significant. This result was surprising because it was hypothesized that tutors' decisions about how to answer tough questions would be related to their motivation to take such risks. Instead, efficacy and interest beliefs did not seem to influence how peer tutors chose to answer questions. One explanation for this result is that question-answering behaviors could be influenced by other factors such as perceptions of the tutoring role. Peer tutors may perceive that they must try to answer every question asked by their tutee if they are able to, regardless of motivation. If this hypothesis is true, then it would suggest that manipulating tutors' role perceptions could have an impact on tutor learning and behavior that is less influenced by motivational concerns. More research would be needed to assess that hypothesis.

Finally, the proportion of monitoring statements that occurred in an elaborative context (i.e. in conjunction with elaboration statements) was significantly and positively correlated with topic interest, $r(33) = .34$, $p = .046$ and tutoring interest, $r(33) = .39$, $p = .021$. These results paralleled the episode-level results regarding sense-making. Tutors who found the subject matter and tutoring task more enjoyable were more likely to integrate their metacognitive and elaborative activities, such as exploring new ideas out of curiosity or to address a misconception.

7.4 SUMMARY

In support of Hypotheses K through N, analyses of peer tutors' efficacy and interest beliefs showed that these motivational factors were an important component of tutor learning. Tutors who felt more confident in their tutoring and learning abilities acquired a stronger factual knowledge and deeper understanding of the domain. Subsequent analyses showed that these factors, perhaps especially tutors' feelings of topic interest (Hypothesis M) and tutoring interest (Hypothesis N), were related to observed tutoring behaviors. Tutors who were more interested engaged more reflective knowledge-building activities, such as elaborated reviewing and sense-making, which contributed to their effective learning of the material.

These analyses also suggested an interesting additional role for prior knowledge. We previously discussed how prior knowledge gave tutors more information to draw upon to give explanations and answer questions. However, their learning behaviors were not entirely constrained by prior knowledge. Tutors with both more and less knowledge coming into the tutoring sessions could still engage in and benefit from knowledge-building activities. What these motivational analyses showed is these behaviors were influenced by peer tutors' attitudes toward the material, which were in turn influenced by their prior knowledge. Thus, peer tutors may have constrained their *own* learning and behaviors based on their *perceptions* of their prior knowledge and the material. "Unmotivated" tutors may choose to ignore or resist the opportunities and support for knowledge-building inherent to the tutoring process.

8.0 DISCUSSION

Prior research has established that students can benefit academically from tutoring their peers, and this learning is inherently supported by nature of the tutoring task. For example, tutors must generate correct and coherent explanations to convey the material, and respond to questions that challenge the tutors' understanding. Such tasks and tutor-tutee interactions present opportunities for tutors to integrate new and prior knowledge, generate new ideas, and reflect upon their own learning (Fuchs et al. 1997; King et al. 1998; Roscoe & Chi, in-press-A). The current study replicated prior results (Roscoe & Chi, in-press-A) to show that peer tutors whose explanations incorporated elaborated reviewing and sense-making, analogous to self-explaining (Chi et al. 1989) obtained a more complete and deeper understanding of the material than tutors whose explanations did not (Hypothesis A). Similarly, I replicated the finding that deep tutee questions elicited deeper and more metacognitive tutor responses (Hypothesis E). Novel analyses conducted in this study further demonstrated that tutors who engaged in deeper question-response interactions gained a better understanding of the material (Hypothesis F).

A key aspect of these learning opportunities is that they were not fully dependent upon tutors' prior knowledge. Certainly, tutors who learned more during the initial studying phase had more knowledge to draw upon during the tutoring phase. Some of these tutors may have simply been good students, who knew how to study text materials more effectively or carefully. As a result, they acquired stronger initial knowledge of the material prior to tutoring. This prior knowledge enabled them to give higher quality knowledge-telling explanations, which included elaborations that tutors had previously generated. Knowledge-building activities such as sense-making were also facilitated by having a stronger foundation. However, both high and low performing tutors were able to engage in and benefit from knowledge-building. Thus, even if students were somewhat limited by prior knowledge or general ability, these individual constraints did not completely determine tutors' behaviors or learning.

This finding is made more salient by comparing the tutors to the reviewers. Active reviewers' knowledge-building behaviors (i.e. concept elaboration, elaborated review, and sense-making) were significantly influenced by prior knowledge. Reviewers who did well in the study learned a lot during the initial studying phase, probably because they spontaneously self-explained. These students then continued to self-explain and improve in the think-aloud phase. In contrast, low performing reviewers likely struggled to learn throughout both phases. Reviewers' learning seemed to be limited by their prior knowledge and reading skills.

These results suggest that the peer tutoring process not only supports tutor learning, but does so in a way that partly transcends tutors' prior knowledge. This effect might be conceptualized in terms of the "zone of proximal development" or ZPD (Vygotsky, 1978). ZPD refers to that range of skills or learning that is just above or out-of-reach learners' current state of understanding. It represents what the learner could achieve if they put forth the effort and received the right kind of help. Expert tutors, intelligent tutoring systems, and peer tutors can provide this kind of help for tutees in the form of timely hints, explanations, and questions that scaffold the tutees' active learning (Foot et al. 1990; Cohen, 1986; VanLehn et al. 2007). If tutees are struggling to understand a formula, a tutor might first explain each variable, and then walk the tutee through an example application of the formula. In this manner, the tutee is supported in using and acquiring knowledge that was too difficult for them to get on their own.

A similar process might work in reverse to benefit the peer tutor. It was argued that peer tutors' knowledge is likely to be flawed because they are still novices in many ways. For tutors who are only moderately more advanced than their tutees, tutee errors and confusions probably have a good chance overlapping with shaky areas of the tutors' own knowledge. The tutoring interaction then provides the tutor a forum for discussing, evaluating, and building upon that knowledge. For example, in the example above, a peer tutor would likely benefit from trying to generate examples and explain the solution steps. Even when tutors do understand the material, they might benefit from justifying their perspective to a more novice learner (Doise, 1990). In this way, tutor-tutee interactions can help to extend peer tutors' learning and strategies beyond what they could have achieved by studying the material on their own.

An important caveat regarding tutor learning, however, is that the richness of the tutoring process can only provide *opportunities* for reflective knowledge-building. There are no guarantees. As hypothesized (Hypotheses B and C), peer tutors showed a strong knowledge-telling bias, exhibiting a high frequency and proportion of knowledge-telling. Shallowly paraphrasing the text materials, or rehearsing the same ideas again and again, contributed little to tutor learning. These activities were not necessarily harmful to the tutors; paraphrasing probably helped tutors put the material into more familiar terms for themselves and their tutee. Reading programs often include paraphrasing as a strategy (e.g. Fuchs, Fuchs, Kazdan, & Allen, 1999; McNamara, 2004). The problem occurred when tutors engaged in knowledge-telling to the exclusion of deeper strategies. As a consequence, their understanding was not much improved, which replicated prior research (Fuchs et al. 1997; King et al. 1998; Roscoe & Chi, in-press-A).

This knowledge-telling bias was not due to the lack awareness of effective strategies (Hypothesis D). Two groups of peer tutors were introduced to knowledge-building strategies developed from prior work (Roscoe & Chi, in-press-A). Both groups were told to use concept elaboration, elaborated reviewing, and sense-making strategies as they taught, which were framed as either explaining strategies or question-answering strategies. This manipulation had little to no effect on tutors' behaviors or learning. Tutors who received the strategies were almost identical to the no-strategy tutors and the reviewers.

8.1 POTENTIAL CAUSES OF THE KNOWLEDGE-TELLING BIAS

In a recent review of the literature, Roscoe & Chi (in-press-B) argued that this widespread knowledge-telling bias contributed to the weak tutor learning gains observed across many programs. They concluded that the knowledge-telling bias, and potential sources of the bias, were a critical challenge and opportunity for future research. Such explorations could provide insights into why peer tutors avoid knowledge-building strategies, and perhaps lead to new training methods designed to overcome such barriers. In the current project, two potential sources of the knowledge-telling bias were evaluated: inadequate metacognitive monitoring and regulation, and negative self-perceptions of ability and interest.

There was little evidence that tutor learning was hampered by a lack of self-monitoring, or a lack of monitoring of key factors such as comprehension or comprehension failures. High and low performing tutors did not differ in monitoring at the episode level, or in the content or valence of monitoring statements. Moreover, comprehension-monitoring statements were the most common type of self-evaluation uttered, and negative and uncertain valence statements occurred more often than positive valence statements. These results indicated that high and low-performing tutors were similarly able to detect and verbalize barriers to their understanding. However, overall monitoring at the episode level, and comprehension-monitoring at the statement level, were positively related to knowledge-building activities, which provided partial support for Hypotheses G and H). Hypotheis I (regarding statement valence) was not supported.

The main metacognitive difference between high and low-performing tutors appeared to be in their self-regulatory behaviors. At a larger grain-size (i.e. episodes), high performing tutors engaged in more knowledge-building activities such as sense-making. By definition, sense-making involved using prior knowledge and logical reasoning to make sense of confusing ideas or to satisfy curiosity. Thus, these episodes represented instances where tutors monitored their own knowledge, observed a gap or deficit, and used knowledge-building to address the problem. At a smaller grain-size (i.e. individual statements), high-performing tutors' monitoring statements were more likely to occur in an elaborative context (Hypothesis J). This meant that they evaluated the quality of their elaborations, and also generated elaborations in response to self-evaluations. Although self-monitoring provided opportunities for knowledge-building, only tutors who took advantage of these opportunities benefited.

An interpretation was that ineffective self-regulation was more of a product of the knowledge-telling bias than a cause of it. Upon encountering an impasse, some tutors responded with knowledge-building while others relied on knowledge-telling strategies. The question still remained why peer tutors made these choices. One metacognitive variable that could not be assessed in the present study was the accuracy of tutors' self-evaluations. There were relatively few positive-valence statements, and so the tutors probably were not overconfident. However, high and low performers may have differed in how well they identified the causes of their confusion. High performers may have been more aware of which concepts that they misunderstood, enabling them to use knowledge-building strategies to address the specific problems. In contrast, low performers may have realized they misunderstood, but had no idea why. Thus, they could only ignore the problem or rely on shallower strategies for repair. This is an interesting issue that deserves further study in future research.

Peer tutors' self-perceptions and attitudes, on the other hand, had a clear relationship to tutor learning and behavior, as hypothesized (Hypotheses K through N). Self-efficacy beliefs were positively linked to learning outcomes and elaborated reviewing. Tutors who felt more capable of learning and teaching the material seemed more willing to engage in knowledge-building, and thus learned more. The data suggested an even stronger impact of tutors' interest in the material and tutoring. Tutors' perceived enjoyment and value of the subject matter and tutoring task were positively and significantly correlated with assessments of factual recall and comprehension. In addition, these variables were associated with a higher proportion of sense-making and monitoring in an elaborative context. Tutors' interest seemed to be an important factor in their decisions about how to regulate perceived errors and comprehension failures, with obvious consequences for their learning of the material.

One reason that interest may have had a greater impact is that interest represents an intrinsic drive to pursue some idea or activity. Peer tutors with low interest, but high self-efficacy, might still rely on knowledge-telling strategies simply because they do not care. In contrast, peer tutors with high interest, but lower self-efficacy, might continue strive to learn and teach as best they can. If tutors' desire to succeed is strong enough, they may expend considerable effort even though the odds seem stacked against them. Such scenarios may be especially relevant when tutors encounter challenges, such as perceived comprehension failures, that require significant mental effort to overcome.

Analyses of motivation also showed that the influences of prior knowledge could operate in indirect ways. Tutors who learned more during the initial studying phase also felt more positive as they went into the tutoring sessions. This made sense; tutors who knew the material better probably felt more empowered to teach it. Tutors who learned less were probably worried, and rightly so, about having to explain the material to a peer. Thus, although prior knowledge did not prevent tutors from engaging in knowledge-building, it indirectly impacted tutor learning and behaviors by affecting their attitudes, which then influenced their decisions about how to tutor.

Future study of motivational issues might be expanded in several ways. One approach would be to examine self-efficacy and interest beliefs in a more fine-grained manner. We conducted principal component analyses in order to validate that our Attitudes Toward the Material Scale (ATMS) and Attitudes Toward Tutoring Scale (ATTS) were separable into our desired scale scores. For our purposes, it was sufficient that the scale items clustered generally into our theory-guided efficacy and interest categories. However, both scales comprised more than two components. Future analyses and research could be used to further explore and test the items underlying the components, and then refine the items and sensitivity of the scales.

Measures to assess tutor motivation could be useful for researchers and teachers. Such assessments might help to identify motivational barriers to effective tutoring before they are harmfully manifested in the tutoring sessions. For peer tutors who lack self-efficacy or interest, educators might intervene by altering the tutoring tasks or curriculum in ways that would better support motivation (see below). Such information might also help with assigning tutor and tutee pairs. One might wish to partner particularly enthusiastic and confident tutors with the tutees who need the most help and require more tutor effort. In research, such tools would provide an additional source of data for analyzing or diagnosing experimental results. Researchers could explore the extent to which high versus low efficacy or interest impacted treatment fidelity, or whether they mediated group differences.

Another way to expand this line of research is to consider other motivational variables. Self-efficacy and interest were chosen because they have often been studied as outcomes of peer tutoring. However, other constructs may certainly be relevant. One interesting area of research is achievement goal theory (e.g. Ames & Archer, 1988; Dweck & Leggett, 1988; Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002; Pintrich, 2003). The fundamental tenet of this theory is that learners' behaviors and outcomes are influenced by their personal reasons for learning. Two broad goal orientations, "mastery" and "performance" goals, are typically described.

Students who adopt mastery goals learn for the sake of developing their knowledge. They are internally motivated and put forth effort above and beyond what is required for the task. As a result, mastery-oriented students typically show higher grades and test scores. Other students are more externally driven. Instead of pursuing knowledge for its own sake, they focus on comparisons to others and external rewards (e.g. grades). They may expend the necessary effort to look good and achieve rewards, or to avoid embarrassment, but may stop once their goal is attained. Such students can be sometimes be successful in settings (e.g. college) where learning tasks and rewards are well-defined.

As with self-efficacy and interest, learning goals might map on to peer tutors' attitudes in several ways. Mastery-oriented tutors might naturally view tutoring a peer as another learning opportunity. They would engage in reflective knowledge-building activities and try to improve their own understanding simply because that is their default inclination. They might hold a similar standard for their tutee. They may to teach not only the basic material, but also try to guide the tutee to think about deeper issues and tackle challenging problems. In contrast, performance-oriented tutors might focus more on displaying their knowledge to the tutee. They may be more prone to lecturing based on their prior knowledge, but not as interested in building upon that knowledge. Moreover, their goals for tutoring may simply to do what is necessary to help the tutee "just pass the test," perhaps by focusing on memorizing key terms or formulas.

In sum, peer tutors' attitudes and self-perceptions can influence their tutoring behaviors and learning, and so it is important to attend to this aspect of the tutoring experience. A lack of interest in the material or tutoring task, and in some cases low self-efficacy, appeared to be plausible sources of the knowledge-telling bias.

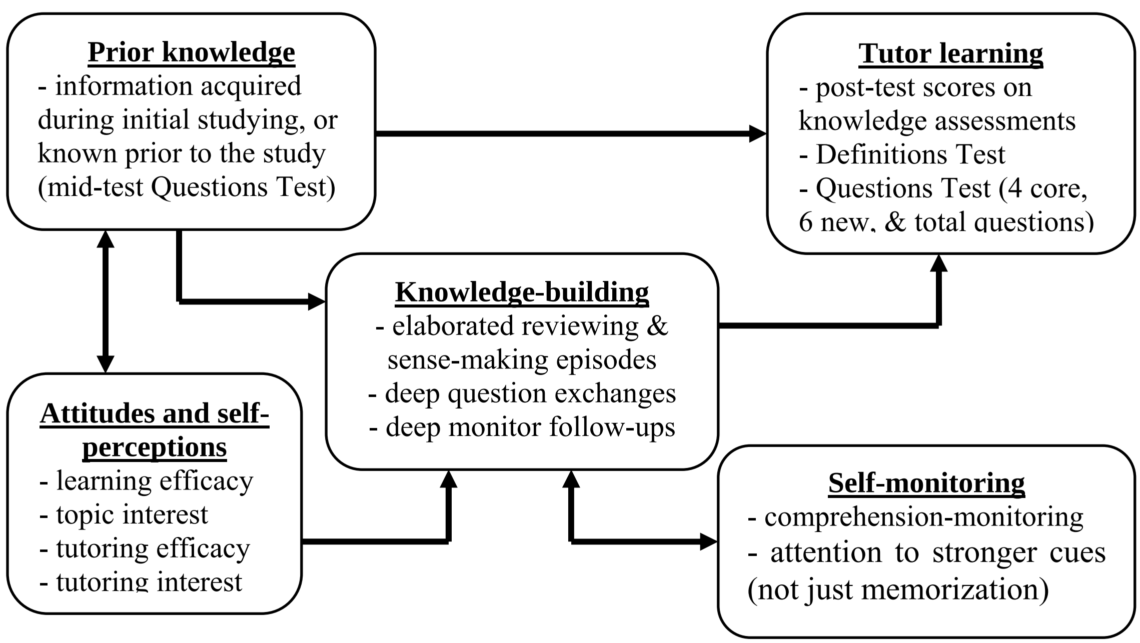

Figure 1. Graphical summary of relationships between learning, behavior, and motivation

Figure 1 summarizes of the patterns of relationships between learning, prior knowledge, knowledge-building, monitoring, and motivation. Unidirectional arrows show relationships with a specific order in time (i.e. knowledge-building occurred before post-tests). Bidirectional arrows indicate relationships between variables assessed simultaneously (i.e. prior knowledge and motivation were assessed at the same time). Thus, prior knowledge was directly related to learning, but also indirectly related to learning via knowledge-building and motivation. Knowledge-building activities were related to learning, and seemed to be influenced by motivation and self-monitoring activities and tutors' attitudes. Prior knowledge and knowledge-building activities were the best predictors of peer tutors' final understanding of the material.

8.2 IMPLICATIONS FOR TUTORING PROGRAM DESIGN

Analyses identified tutors' negative attitudes as possible causes of the knowledge-telling bias. The implication of these findings is that peer tutoring programs should incorporate training activities and environments that support the development of positive tutor interest and engagement. In the following sections I draw upon the principles of self-determination theory (e.g. Deci, 1992; Ryan & Deci, 2000) to begin a discussion of interest-supporting factors relevant to peer tutoring. Future research will be needed to empirically refine these initial suggestions.

Self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000) describes the environmental and personal factors that support or undermine interest and engagement. In this theory, interest is linked to the concept of "intrinsic motivation," which stems from personal desires and preferences. People engage in intrinsically-motivated behaviors because they find the activities inherently rewarding and enjoyable. Such behaviors are freely and spontaneously chosen, and characterized by deeper engagement and effort. In contrast, extrinsic motivation is externally-regulated. People engage in extrinsically-motivated behaviors primarily to attain rewards, such as money or good grades, or because they are compelled to do so by work obligations, peer pressure, etc. These behaviors receive much less effort and engagement.

Three "needs" have emerged as important prerequisites of interest: competence, relatedness, and autonomy. When these needs are supported, learners are more likely to develop interest in the topic or task (Deci et al. 1996). However, in cases where learners might otherwise show strong interest, these feelings may be thwarted by a lack of competence, relatedness, or autonomy. To foster interest, tutoring programs might seek to integrate support for these needs.

8.2.1 Supporting competence

Learners need to feel competent and capable of engaging in academic tasks successfully (i.e. self-efficacy), and need opportunities for self-improvement. Ideal learning environments should provide learners with "optimal challenges" (Deci et al. 1996); learning tasks should be a bit beyond the learners' current knowledge and capabilities, but fully attainable with some effort. Intrinsic motivation is thwarted by challenges that seem impossible (or too easy).

In some ways, peer tutoring may inherently support the need for competence. Tutor-tutee interactions enable tutors to engage in learning behaviors that are just above their current level of ability. Further support could arise from the training process. Exposure to effective tutoring techniques, and time to practice them, could enhance self-efficacy. Practice sessions provide tutors with the opportunity to experience gradually increasing success in using the strategies. If teachers or supervisors are present, they can reinforce these successes through feedback. They can help tutors to appreciate when they have generated knowledge-building elaborations, and help tutors to better recognize and take advantage of knowledge-building opportunities. Thus, some programs may be effective because they not only teach students to engage in reflective knowledge-building, but simultaneously build tutors' efficacy for engaging those behaviors.

The recommendation to provide tutors with knowledge-building strategy training and practice may seem obvious. The truth of the matter is that many tutors do not receive such training, or any training (Graesser et al. 1995; Rohrbeck et al. 2003; Roscoe & Chi, in-press-A; Topping, 1996). Peer tutors, especially those in high school and higher education, are often recruited based on strong course grades. The assumption is that these tutors know the material and so can share it with another student. Once these students begin tutoring, they probably find it much harder than they expected. As a result, they may experience a significant drop in self-efficacy and intrinsic motivation. Providing these tutors with concrete knowledge-building strategies may support intrinsic motivation, tutor learning, and tutee learning.

The need for competence may also be supported by giving tutors opportunities to further develop their knowledge prior to tutoring. Peer tutors, although often more advanced than their tutees, probably still have many knowledge gaps and misconceptions. The perception of inadequate prior knowledge could lead some tutors to doubt their ability to tutor effectively. Opportunities to "prepare to teach" can directly improve the tutors' understanding (e.g. Annis, 1983; Fantuzzo et al. 1989), which should help tutors to reinforce or reaffirm their expertise. Such activities can be combined with tutor training. That is, tutors could review the subject matter while practicing the knowledge-building strategies they have been taught. Of course, giving tutors opportunities to strengthen their prior knowledge would also directly enhance their ability to engage in reflective knowledge-building activities. Thus, preparation-to-teach might simultaneously address both motivational and knowledge barriers to tutor learning.

8.2.2 Supporting relatedness

Learners need the material to be personally relevant, and for learning tasks to provide meaningful connections with others. Relatedness involves learning topics that overlap with the learners' hobbies and interests (e.g. sports), and also stems from opportunities to use prior knowledge in ways that makes the person feel valued. Relatedness is also supported by allowing learners to interact more closely with respected adults, and have social and learning interactions with peers. Intrinsic motivation is hampered by feelings of isolation or tasks that seem arbitrary.

Some aspects of peer tutoring naturally address the need for relatedness. Tutoring a peer provides an academically and socially valued service to the community, which may improve tutors' self-image (Allen, 1983; Cohen, 1986). The tutors are using their prior knowledge to help another person, and often receive praise and gratitude for doing so. In addition, tutors get to interact with other students. If the tutor and tutee are close in age, then they might develop a friendship that extends past the tutoring environment. If the age gap is larger, they might develop a relationship similar to older and younger siblings (e.g. Juel, 1996).

In order to support relatedness, peer tutors' preparation activities might be combined with opportunities to learn more about their tutees' and the curriculum. When they are refreshing their prior knowledge, tutors might be given copies of their tutees' textbooks and class materials. If feasible, peer tutors could be invited to attend some of the tutees' classes. In this way, the tutors might better appreciate the learning challenges faced by their pupil, and the importance of providing tutoring help. Similarly, tutors might benefit from meeting with tutees' classroom teachers. Teachers can provide insights about the material and teaching, and the interaction would give tutors with the chance to interact with an authority figure as a "colleague."

Another way to support relatedness may be to foster a sense of community among the tutors. When tutoring occurs in a classroom, this community may already exist. But after-school or out-of-school tutoring may not support interaction between tutors if they are on different schedules (i.e. they tutor on different days or in different locations). Periodic meetings of the tutors could serve several purposes. Social interaction among tutors would provide opportunities to meet new people and perhaps make new friends. These gatherings could also act as a forum for peer tutors to share and reflect upon their strategies, successes, and setbacks. Tutors may be able to learn new techniques from each other, and receive suggestions for overcoming difficulties. Such forums might support both competence and relatedness needs.

8.2.3 Support for autonomy

Learners need to feel that they have control over their behaviors, and that their feelings are taken into account in any decisions regarding themselves. Autonomy-supportive teachers and settings provide learners with choices, acknowledge students' opinions (including criticisms), and avoid controlling strategies (Reeve, Jang, Carrell, Jeon, & Barch, 2004). Controlling strategies try to force or coerce students to follow an agenda that "defines what students should think, feel, and, do" (p. 148). This includes giving rewards or punishments tied to specific behaviors, or using "pressuring language" that commands students how to behave. Intrinsic motivation is hindered in settings where external regulations seem to strip away students' choices and preferences.

Unlike competence and relatedness, the need for autonomy may not be well supported by current tutor training methods. When tutors do not receive strategy training, learning and self-efficacy may be hindered. However, extensive tutor training may thwart the need for autonomy because the strategies are often highly specified, and there is pressure to maintain high treatment fidelity (i.e. for tutors to follow the instructions carefully). As a result, peer tutors may be told repeatedly what, when, and how to tutor. These problems could be exacerbated when tutors are offered rewards for using the strategies. External rewards are a form of extrinsic-regulation, which undermines learners' intrinsic motivations for engaging in tutoring (Deci, Koestner, & Ryan, 1999; Lepper & Hodell, 1989) Tutors may become less interested in learning the material and tutoring, and thus simply "go through the motions" until they attain the rewards.

How can one balance the need for training with the potential danger of thwarting tutors' autonomy? One method may be to provide tutors with meaningful strategy choices. Research on choice has shown that giving students options in topics and tasks can increase interest and engagement (Black & Deci, 2000; Cordova & Lepper, 1996; Schraw, Flowerday, & Reisetter, 1998). Typically, peer tutors will not have many options in terms of the subject matter, because curriculum issues are dictated primarily by the needs of the tutee. However, greater flexibility is feasible with tutoring techniques. Tutors are typically introduced to a single set of strategies, and instructed to use the strategies as directed, and in a particular order. A modification of this approach would be to introduce tutors to multiple strategies that they could choose from.

For example, tutors might be taught strategies in which they use question-stems to create deep questions (King et al. 1998), give conceptual explanations when tutees make mistakes (Fuchs et al. 1997), and sense-making to revise their own knowledge (Roscoe & Chi, in-press-A). Although tutors might learn them all, during tutoring they would be allowed to use whichever strategies they preferred. If they felt that one strategy was more suitable or "more fun," then they could switch. Because all of these strategies involve knowledge-building, tutor learning would be supported in each case. This freedom might lead tutors to combine multiple techniques, and knowing more strategies might further support tutors' sense of competence.

Another way to support autonomy is for program supervisors to solicit and listen to peer tutors' complaints or suggestions about tutoring difficulties. If tutors feel that tutoring is boring, unimportant, or ineffective, it is important for educators to attend to these concerns. If tutors' negative responses are ignored, then their intrinsic motivation will be hindered. Supervisors may also benefit because tutors' complaints can reveal ways to alter or improve the program. Sometimes students will tell you exactly why they are not following instructions.

8.3 CONCLUSION

Research on peer tutoring has demonstrated the potential for this activity to support impressive learning outcomes for student tutors. The knowledge-telling bias, however, presents a significant obstacle to actualizing these benefits. The current project has contributed to a new research approach that seeks to understand the causes of this bias. Instead of asking "What do tutors need to do to learn?" I have asked "Why don't peer tutors engage in the behaviors we know are beneficial?" This reformulation opens up new kinds of research on peer tutoring and learning. An initial exploration into these issues suggested that tutors' attitudes, often assessed as outcomes of tutoring, were also important predictors of learning and behavior. Future research should continue to explore these ideas, and other potential causes of the knowledge-telling bias. Metacognitive and motivational barriers are just two possible variables; the current project also highlighted the importance of prior knowledge. Although not assessed here, tutors' perceptions of tutoring could also be influential. Identifying and teasing apart the various sources of the knowledge-telling bias could ultimately lead to innovative changes across a broad array of tutoring programs, which will increase the scope and magnitude of the tutor learning effect.

APPENDIX A

MATERIALS AND MEASURES

A.1 PARTICIPANT BACKGROUND QUESTIONNAIRE

AGE: **YEAR IN SCHOOL:** **SAT SCORES:** **V:** **M:**

GENDER: **CURRENT GPA:**

RACE/ETHNICITY: **MAJOR/MINOR:**

1. Have you ever provided **formal** tutoring to another student? (Formal tutoring means that you participated in an established tutoring program and/or were paid to tutor.)

YES… I have provided formal tutoring **NO… I have not provided formal tutoring**

About how many hours of **formal** tutoring have you provided in the last three years? _________

2. Have you ever provided **informal** tutoring to another student? (Informal tutoring refers to tutoring a friend, lab partner, or classmate outside of an established tutoring program and without any payment.)

YES… I have provided informal tutoring **NO… I have not provided informal tutoring**

About how many hours of **informal** tutoring have you provided in the last three years? _________

3. Have you ever taught in a **classroom** setting? (Examples include giving a classroom lecture or leading group discussions as part of a class assignment.)

YES… I have taught in a classroom **NO… I have not taught in a classroom**

4. Please describe your past tutoring or teaching experiences (if any).

5. What are three things that tutors/teachers can do to help their students learn effectively?

6. What are three things that tutors/teachers might do that *would not* help their students learn effectively?

A.2 PARTICIPANT FEEDBACK QUESTIONNAIRE

1. How would you describe the difficulty of the content taught in this study? (please circle one response)

Much harder than content I am used in to in my classes	Somewhat harder content than I am used to	The same as content I am used to in my classes	Somewhat easier content than I am used to	Much easier than content I am used to in my classes

Optional comments:

2. How would you describe the quality of the text and diagrams used in this study? (circle one response)

Very poor clarity and organization	Poor clarity and organization	Average clarity and organization	Good clarity and organization	Very good clarity and organization

Optional comments:

3. If you have previously taught or tutored another student in a one-on-one setting, please indicate how similar this study was to your previous teaching/tutoring experience. (circle one response)

Not at all similar to my past experience with tutoring	Somewhat dissimilar to my past experiences	About equal similarities and differences	Somewhat similar to my past experiences	Very similar to my past experiences with tutoring

Optional comments:

4. Describe the strategies that you used **during the preparation phase** (the first studying session) at the beginning of the study.

5. Describe the strategies that **you** used in your assigned learning task **during the reviewing/tutoring phase**. Please describe the strategies that you **actually** used, even if you chose to do something different than the assigned strategies that you were given.

6. Describe the strategies that **your student** or **your tutor** used in your assigned learning task **during the tutoring phase**. If you did not have a partner, you can ignore this question.

Please write any other additional comments you have about this study on the back of this sheet.

A.3 HUMAN EYE AND RETINA TEXT

The eye is the sensory organ that enables human vision. Humans have two eyes located at the front of the head. Each eye is about 1 inch in diameter and surrounded by an opaque, protective layer called the sclera.

Light entering the eye first passes through the cornea, which is a clear membrane covering the front of the eye. The shape of the cornea is responsible for about 70% of the focusing power of the eye. It must be transparent and undamaged in order for a clear image to be formed on the retina.

Behind the cornea is the iris, which contains smooth muscle fibers arranged in circular and radial directions. The iris opens and closes to adjust the amount of light entering the eye. Pigments in the iris are responsible for the visible color of the eye.

The pupil is the opening in the center of the iris. Light exiting the cornea passes through this opening. The size of the pupil is decreased when the circular muscles of the iris contract. The pupil becomes dilated when the radial muscles contract.

The cornea does most of the focusing of light, but additional focusing occurs by varying the thickness of the lens. This process is called accommodation. In this process, the ciliary muscles contract, which increases the curvature and focusing power of the lens. This process allows us to keep an image on the retina clear as we view things at varying distances.

After light exits the lens, it passes through the vitreous humor, which is a clear, jelly-like substance in the middle of the eye. One purpose of this substance is to maintain the shape of the eye. In addition, the vitreous humor has a refractive index similar to that of the lens, which prevents further bending of the light.

After light passes through the vitreous humor it reaches the retina. The retina covers an area of nearly 200 degrees (as measured on the circumference of a circle) on the inner surface of the eye. Photoreceptors are located on the back of the retina. Neurons and blood vessels are located in front of the receptors.

Photoreceptors convert light energy into neural energy through the process of transduction. Photoreceptors contain molecules called photopigments that can absorb light photons. When this occurs, the pigments undergo a molecular change that can produce a neural signal.

There are two kinds of photoreceptors: rods and cones. Rods are the rod-shaped receptors that are responsible for vision in low levels of light. However, although the rod system is sensitive in the dark, it cannot resolve fine details. There are about 120 million rods in the retina.

Cones are the cone-shaped receptors that are responsible for vision in high levels of light. The cone system is capable of resolving fine details, and also enables color vision. There are about 7 million cones in the retina.

The fovea is the most sensitive section of the retina for detecting patterns of light. It covers an area of about 1 degree. The fovea contains most of the cones receptors, which decrease in number dramatically with distance from the fovea.

There are no rods in the center of the fovea. Rods are most prevalent at a distance of 16 degrees from the fovea. The number of rods decreases gradually out to about 100 degrees at the edge of the retina.

Neural signals from the receptors travel to the brain via synaptic connections. The photoreceptors synapse onto bipolar cells. In the fovea, cones connect to bipolar cells in a one-to-one ratio. In the periphery, the ratio increases to up to 100 rods per bipolar cell.

The high convergence of receptors in the periphery means that bipolar cells in this region cannot determine which of their receptors have been stimulated by light. However, the low degree of pooling in the fovea leads to a higher specificity of excitation in that area.

Each bipolar cell is connected to a ganglion cell through a second synapse. Ganglion cells respond to complex patterns of light and dark, and also react to colors. There are less than 1 million ganglion cells. In the fovea, each bipolar cell generally connects to one ganglion cell, whereas in the periphery, many bipolar cells converge on a single ganglion cell.

Ganglion cells have elongated axons that form the fibers of the optic nerve. The location where the optic nerve exits the eye is called the blind spot. There are no receptors located in the blind spot. The optic nerve carries signals to the brain, which then combines the signals to form the image that we perceive.

There are two other types of neural cells found in the retina. Horizontal cells connect receptors to other receptors. Amacrine cells connect ganglion cells to other ganglion cells, and bipolar cells to other bipolar cells.

Horizontal and amacrine cells do not send signals directly to the brain, but instead send and modulate signals between other cells in the retina. As a result, neural signals in the retina flow both directly toward the brain, and laterally across the retina before going to the brain.

A.4 DEFINITIONS TEST

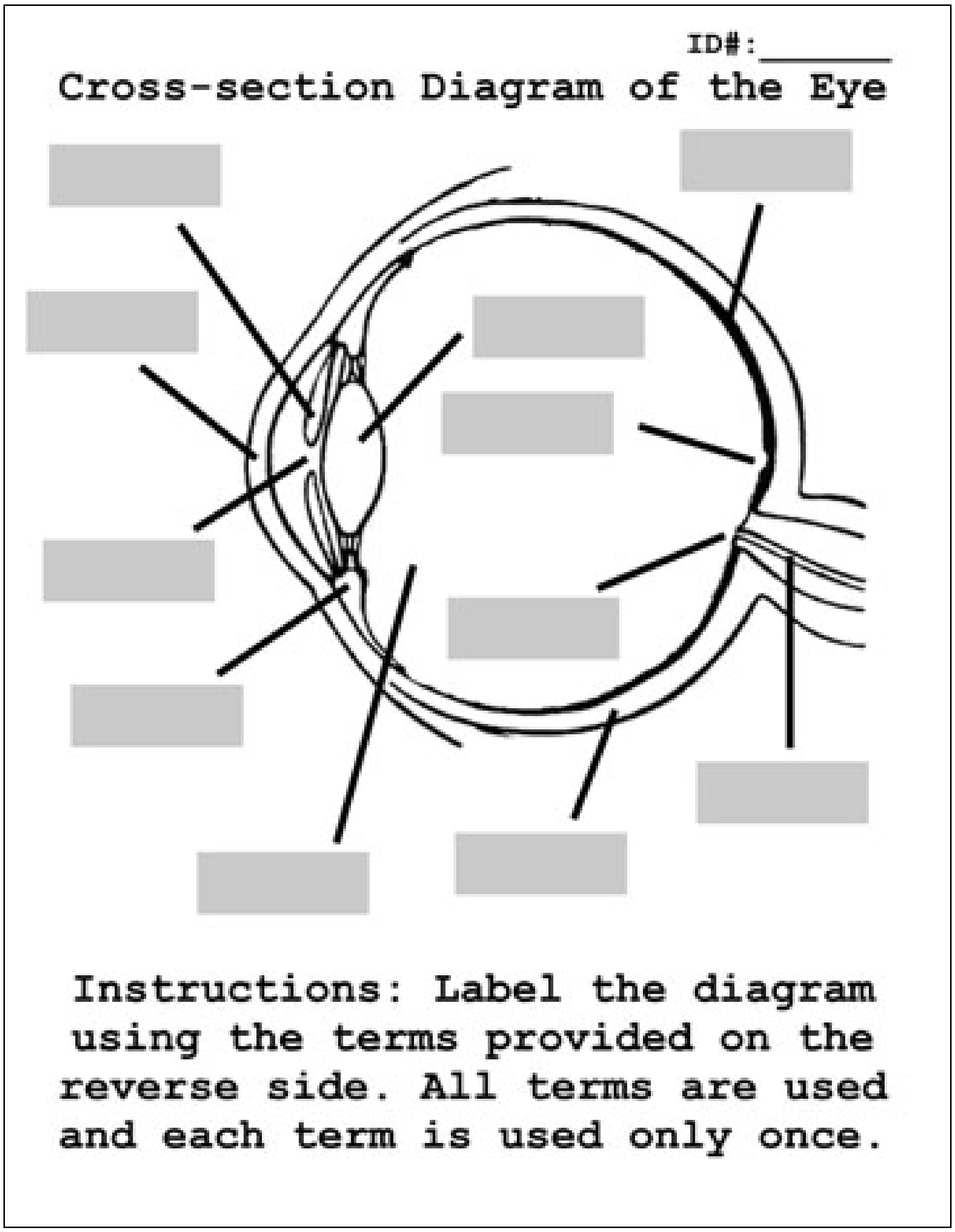

Figure 2. Definitions Test: Unlabeled cross-section diagram of the eye

Cross-section Diagram of the Eye

Instructions: Define the terms below. Your definitions should include information about the structure and function of the component, along with any other facts you feel are relevant.

Blind spot:

Ciliary muscles:

Cornea:

Iris:

Fovea:

Lens:

Optic nerve:

Pupil:

Retina:

Sclera:

Vitreous humor:

Figure 3. Definitions Test: Eye diagram terminology

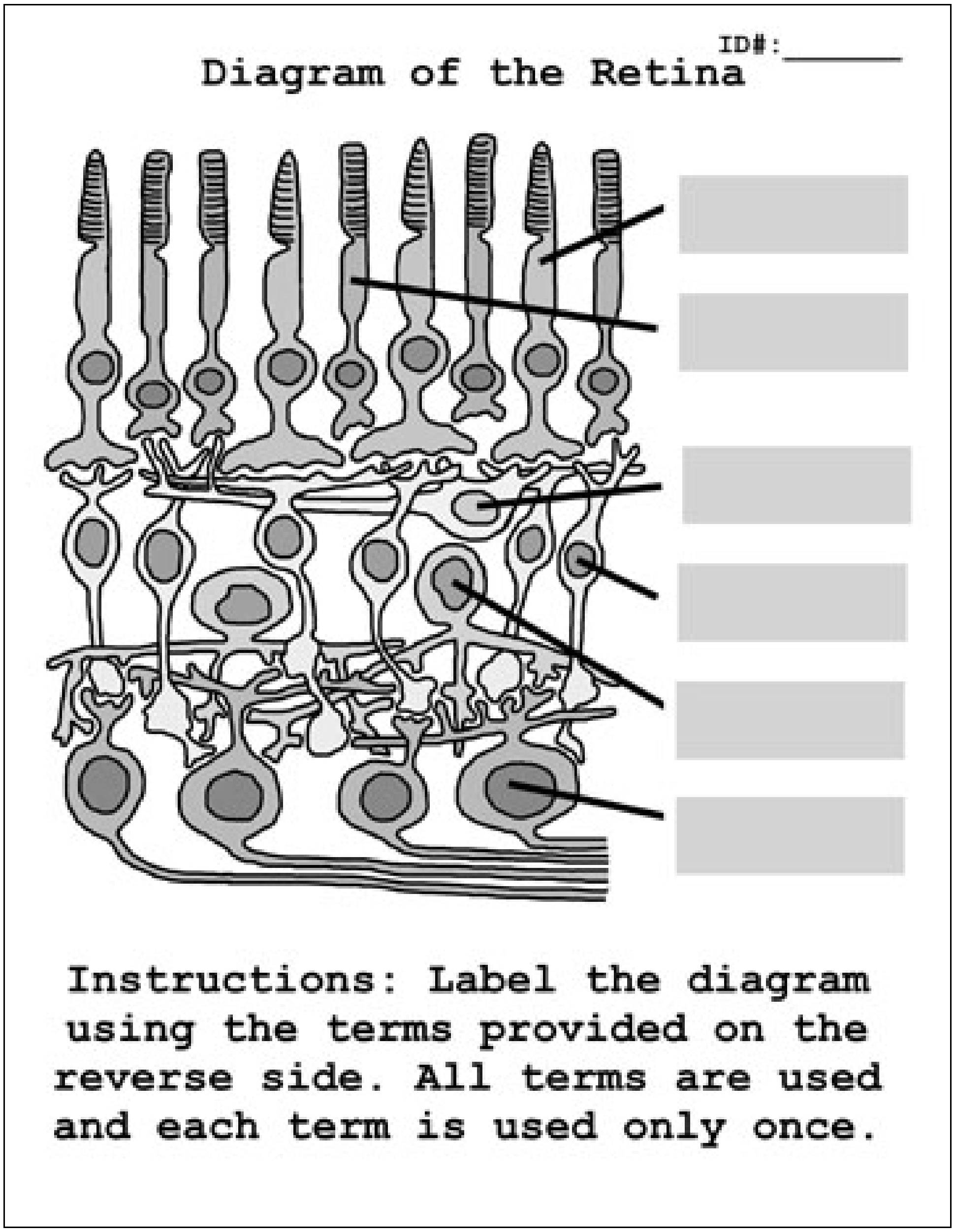

Figure 4. Definitions Test: Unlabeled drawing of the retina

Diagram of the Retina

Instructions: Define the terms below. Your definitions should include information about the structure and function of the component, along with any other facts you feel are relevant.

Amacrine cell:

Bipolar cell:

Cone receptor:

Ganglion cell:

Horizontal cell:

Rod Receptor:

Figure 5. Definitions Test: Retina diagram terminology

A.5 QUESTIONS TEST

Instructions:

The following questions are designed to test your deeper understanding of the material. These questions test your ability to explain how the eye works, integrate your knowledge of eye components, and apply your knowledge to new scenarios. There are 10 questions.

Please read each question carefully. Some questions have multiple parts. Some are new questions that you have not seen before.

Please answer each question as completely and accurately as possible. You may include diagrams as part of your answer. Please write clearly.

(1) What components of the eye are involved in focusing and directing light towards the retina? How do these components accomplish this process? (core question)

(2) How is light energy entering the eye converted to neural signals in the retina? Why is this process necessary? (core question)

(3) What are the most direct and least direct pathways that neural signals can follow to reach the brain from the photoreceptors? (core question)

(4) Why is the fovea the most sensitive section of the retina for detecting light patterns? Why is vision in the periphery less clear? (core question)

(5) How is the amount of light entering the eye regulated by the muscles of the iris? Why is this function important for normal vision? (new question)

(6) What are the two main functions of the vitreous humor? What properties of the vitreous humor are responsible for these functions? Why are these functions important for normal vision? (new question)

(7) What are the major functions of the cone receptor system and the rod receptor system? What causes these systems to possess these properties? (new question)

(8) Some animals can see in the dark. What are several ways that animal eyes might be structured differently than human eyes so that this is possible? (new question)

(9) Eyeglasses may be worn to correct for several visual impairments. What underlying problems may cause a person to need glasses? How do glasses compensate for impaired eye functions? (new question)

(10) Colorblindness is a disorder in which a person is unable to see colors. What are several potential problems underlying this disorder? How could a person be “blind” to only one or two colors? (new question)

A.6 STUDENT SURVEY

Instructions: This survey contains two sets of statements that may describe your attitudes about topics covered in this study and teaching others.

There are no right or wrong answers. We are interested in your *actual feelings*, so it is very important that you respond *honestly* to each item.

You will use the same scale to rate every item:

The 7-point scale:

1 = Strongly disagree. The statement is completely wrong about your attitude.

2 = Disagree. The statement is mostly wrong about your attitude.

3 = Somewhat disagree. The statement is partly wrong about your attitude.

4 = Neither disagree nor agree. The statement is about half-correct and half-wrong about your attitude.

5 = Somewhat agree. The statement is partly correct about your attitude.

6 = Agree. The statement is mostly correct about your attitude.

7 = Strongly agree. It is completely correct about your attitude.

Part One (Attitudes Toward the Material Scale)

Instructions: For the following statements, think carefully about *how you feel about the information you are learning in this study*.

Rate your agreement with each statement on the 7-point scale from "1" (strongly disagree) to "7" (strongly agree) in the gray boxes on the right-hand side.

I know my study skills are good enough that I can learn any topic presented in this study.	

I think that what we are learning in this study is interesting.	

I am sure that I will do really bad on the assessment tests used in this study.	
I rarely choose to engage in activities like this study where I have to learn a lot of new info.	
I feel that I know a great deal about the material covered in this study.	
I think that the information I am learning in this study is useful for me to know.	
My learning strategies are probably not sufficient to understand the topics covered in this study.	
I do not enjoy the topic we are learning about in this study.	
I know that I am capable of comprehending the ideas covered in this study.	
Understanding the topic covered in this study is important to me.	
Compared to other students in this study, I probably know very little about the topic.	
I doubt I will ever use the information in this study in my classes or daily life.	
Compared to other students in this study, I expect to get a high score on the assessment tests.	
I enjoy challenging experiences like this study where I get to learn new things.	
I am certain that I cannot understand the topics presented in this study.	
It is not at all important to me to learn about the topics presented in this study.	

Part Two (Attitudes Toward Tutoring Scale)

Instructions: For the following statements, think about *how you feel about teaching the information in this study to another undergraduate student.*

Rate your agreement with each statement on the 7-point scale from “1” (strongly disagree) to “7” (strongly agree) in the gray boxes on the right-hand side.

I could help students who did poorly on the assessment tests to improve their scores.	
I really enjoy helping others to learn new ideas.	
If the other student became really confused, I would have a hard time figuring out why.	
I do not really care whether I am a good tutor or not.	
I think I could really help another student to think critically about the material.	
It is fun to help other students to think really deeply about what they are learning.	
I am very limited in my ability to help a student understand the topics in this study.	
I really dislike having to explain complicated ideas to other people.	
I know I could create great explanations or examples to teach this information well.	
Helping others to really understand what they are learning is a worthwhile activity.	
I doubt I could help another student to think creatively about the topics in this study.	
It is unpleasant to try to explain ideas to someone who is really confused.	
I know I could think of good questions to test whether the other student understood.	
It is important to me to do a really good job when I teach another person.	
I doubt that I could adjust the level of my explanations to match the student's needs.	
I think that it is fairly useless to try to teach other people new knowledge.	

APPENDIX B

PEER TUTORING AND THINK-ALOUD INSTRUCTIONS

B.1 GENERAL TUTORING INSTRUCTIONS

In the second phase, you will have 20 minutes to use the text, diagrams, and your notes in order to teach this material to your tutee, and try to improve their understanding and test answers as well as your own understanding and test answers. During this phase, you will use a tutoring technique called "Strategic Tutoring." Prior research has shown that "Strategic Tutoring" is a highly effective way to tutor another person, but only if tutors use the strategy properly. I am now going to describe three things you can do to make this strategy work well for you.

First, it is important to cover all of the information in the text. You should teach your student about every major component of the eye and retina, and how they work together. Second, it is important to give high-quality explanations so that your student will really understand the material. Good explanations go beyond what the text says. Third, it is important to respond to your student's questions. You should encourage your student to ask lots of questions, and you should try to answer these questions to the best of your ability.

B.2 EXPLANATION TUTORING INSTRUCTIONS

In the second phase, you will have 20 minutes to use the text, diagrams, and your notes in order to teach this material to your tutee, and try to improve their understanding and test answers as well as your own understanding and test answers. During this phase, you will use a tutoring technique called "Elaborated Explaining." Prior research has shown that "Elaborated Explaining" is a highly effective way to tutor another person, but only if tutors use the strategy properly. I am going to describe three things you can do to make this strategy work well for you:

As the tutor, your job is to explain all of the material and answer any questions your student asks. First, it is important to give detailed and elaborated definitions. Good definitions not only provide the basic information… but also contain new examples and explanations that go beyond what the text says. The best explanations also show how each major concept is related to other major concepts.

Second, after you have explained a few major concepts, it is important to give detailed and elaborated reviews of these ideas to make sure your student understands. Good reviews don't just repeat the exact same information, but also contain new examples and explanations that help the student to understand even better. The best explanations also show how the major concepts fit together into a clear "big picture."

Third, sometimes while explaining you will realize that you don't understand a concept as well as you thought you did. Instead of skimming over what you don't understand, you should attack it head on and try to make sense of the confusing ideas. You should use every resource available (text, diagrams, discussion) to work out the correct explanation for yourself and your student.

B.3 QUESTION TUTORING INSTRUCTIONS

In the second phase, you will have 20 minutes to use the text, diagrams, and your notes in order to teach this material to your tutee, and try to improve their understanding and test answers as well as your own understanding and test answers. During this phase, you will use a tutoring technique called "Elaborated Question-Answering." Prior research has shown that "Elaborated Question-Answering" is a highly effective way to tutor another person, but only if tutors use the strategy properly. I am now going to describe three things you can do to make this strategy work well for you:

As the tutor, your job is to explain all of the material and answer any questions your student asks. The first thing to remember is that every time you answer a question, it is important to give detailed and elaborated definitions. Good definitions not only provide the basic information… but also contain new examples and explanations that go beyond what the text says. The best answers also show how each major concept is related to other major concepts.

Second, whenever your student asks you to go over a concept again, it is important to give detailed and elaborated reviews of these ideas to make sure your student understands. Good reviews don't just repeat the exact same information, but also contain new examples and explanations that help the student to understand even better. The best answers also show how the major concepts fit together into a clear "big picture"

Third, sometimes your student will ask a question, and you realize you don't understand a concept as well as you thought you did. Instead of skimming over what you don't understand, you should attack it head on and try to make sense of the confusing ideas. You should use every resource available (text, diagrams, discussion) to work out the correct answer to the question for yourself and your student.

B.4 THINK-ALOUD REVIEWING INSTRUCTIONS

In the second phase, you will have 20 minutes to review the text, diagrams, and your notes in order to improve your understanding and test answers. During this phase, you will use a learning technique called “thinking out loud.” You will not be using this strategy in the first phase. Prior research has shown that “thinking out loud” is a highly effective way to study, but only if students use the strategy properly. I am now going to describe three things you can do to make this strategy work well for you:

First, it is important to say out loud everything that you are thinking and feeling about the material. You should try to summarize the material to yourself out loud, or explain the material to yourself out loud, or ask yourself questions and try to answer them out loud.

Second, it is important to think out loud for every major concept. You should use the strategy at least once or more for every page of text. The technique will be more effective the more you use it. However, you should remember that this strategy is not called reading out loud. Go ahead and read bits of the text out loud if you want to, but it is much more important for you to think out loud by summarizing, explaining, and asking yourself questions.

Finally, it is important to keep thinking out loud the entire time. As you study the text and your notes, it is easy to forget to keep thinking out loud. In order for this technique to work, however, you should never be silent for more than a few seconds. You need to keep using the strategy continuously. Don’t worry about planning what you are going to say, just let your thoughts flow!

BIBLIOGRAPHY

Ainsworth, S. (1999). The functions of multiple representations. *Computers and Education*, 33, 131-152.

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

Ames, C. & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260-267.

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. & Cromley, J. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523-535.

Azevedo, R., Cromley, J. & Siebert, D. (2004). Does adaptive scaffolding facilitate students ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344-370.

Azevedo, R., Guthrie, J., & Siebert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research, 30*(1&2), 87-111.

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

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148.

Bargh, J. & 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.

Black, A. & Deci, E. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education, 84*, 740-756.

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.

Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist, 34*(2), 75-85.

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.

Chan, C., Burtis, J., Bereiter, C. (1997). Knowledge-building as a mediator of conflict in conceptual change. *Cognition and Instruction, 15*(1), 1-40.

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. & Roscoe, R. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change* (pp. 3-27). Kluwer: London.

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

Cloward, R. (1967). Studies in tutoring. The Journal of Experimental Education, 36(1), 14-25.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155-159.

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

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

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

Cook, S., Scruggs, T., Mastropieri, M., & Casto, G. (1986). Handicapped students as tutors. *Journal of Special Education*, 19(4), 483-492.

Cordova, D. & Lepper, M. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715-730.

Deci, E. L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development*, pp. 43-70. Hillsdale, NJ: Erlbaum.

Deci, R., Koestner, R., & Ryan, R. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627-668.

Deci., R. & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behavior*. NY: Plenum.

Deci, E. L. & Ryan, R. M., & Williams, G. C. (1996). Need satisfaction and the self-regulation of learning. *Learning and Individual Differences, 8*(3), 165-183.

Doise, W. (1990). The development of individual competencies through social interaction. In H. Foot, M. Morgan, & R. Shute (Eds.), *Children Helping Children* (pp. 43-64). NY: John Wiley.

Dufrene, B., Noell, G., Gilbertson, D., & Duhan, G. (2005). Monitoring implementation of reciprocal peer tutoring: Identifying and intervening with students who do not maintain accurate implementation. *School Psychology Review*, *34*(1), 74-86.

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

Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.

Fantuzzo, J., King, J., & Heller, L. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A componential analysis. *Journal of Educational Psychology*, *84*, 331-339.

Fantuzzo, J., Riggio, R., Connelly, S., & Dimeff, L. (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.

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.

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.

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

Fuchs, L., Fuchs, D., Bentz, J., Phillips, N., & Hamlett, C. (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., Fuchs, D., Kazdan, S. & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help-giving. *Elementary School Journal, 99*(3), 201-219.

Gabarino, J. (1975). The impact of anticipated reward upon cross-age tutoring. *Journal of Personality and Social Psychology, 32*(3), 421-428.

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

Ginsburg-Block, M., Rohrbeck, C., & Fantuzzo, J. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology,98*(4), 732-749.

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., Carta, J., & Hall, R. (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.

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*(3), 638-645.

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: Erlbaum.

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.

Koriat, A. & Bjork, R. (2005). Illusions of competence in monitoring one's own knowledge during study. *Journal of Experimental Psychology, 31*(2), 187-194.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

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.

Lepper, M., Drake, M., & O'Donnell-Johnson (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues* (pp. 108-144). Cambridge, MA: Brookline.

Lepper, M. R. & Hodell, M. (1989). Intrinsic motivation in the classroom. In C. Ames & R. Ames (Eds.), *Research on motivation in education (Vol. 3): Goals and cognitions*, pp. 73-105. New York: Academic Press.

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.

Mathes, P. & Fuchs, L. (1994). The efficacy of peer tutoring in reading strategies for students with mild disabilities: A best-evidence synthesis. *School Psychology Review, 23*(1), 59-80.

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

McMaster, K., Fuchs, D., & Fuchs, L. (2006). Research on peer-assisted learning strategies: The promise and limitation of peer-mediated instruction. *Reading and Research Quarterly, 22*, 5-25.

McNamara, D. (2004). SERT: Self-explanation reading training. *Discourse Processes. 38*(1), 1-30.

McNamara, D., Louwerse, M., Cai, & Graesser, A. (2005, January 1). Coh-Metrix version 1.4. Retrieved July 13, 2007, from http://cohmetrix.memphis.edu.

Medway, F. & Baron, R. (1977). Locus of control and tutors' instructional style. *Contemporary Educational Psychology, 2*, 298-310.

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.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.

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

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667-686.

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

Pintrich, P. R. & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

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.

Reeve, J., Jang, H., Carrell, D., Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28*(2), 147-169.

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

Renkl, A. (1997). Learning from worked out examples: A study on individual differences. *Cognitive Science*, 21(1), 1-29.

Rittschof, K. A. & Griffin, B. W. (2001). Reciprocal peer tutoring: Re-examining the value of a cooperative learning technique to college students and instructors. *Educational Psychology, 21*(3), 313-319.

Robinson, D., Schofield, J., & Steers-Wentzell, K. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17*(4), 327-362.

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

Roscoe, R. & Chi, M. (in press-A). Tutor learning: The role of explaining and responding to questions. *Instructional Science*.

Roscoe, R. & Chi, M. (in press-B). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*.

Ryan, R. & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development*, pp. 183-212. Hillsdale, NJ: Erlbaum.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*, 113-125.

Schraw, G., Flowerday, T., & Reisetter, M. (1998). The role of choice in reviewer engagement. *Journal of Educational Psychology*, *90*(4), 705-714.

Schunk, D. H. (1995). Inherent details of self-regulated learning include student perceptions. *Educational Psychologist, 30*(4), 213-216.

Shah, F., Evens, M., Michael, J., & Rovick, A. (2002). Classifying student initiatives and tutor responses in human keyboard-to-keyboard tutoring sessions. *Discourse Processes, 33*(1), 23-52.

Sharpley, A., Irvine, J., & Sharpley, C. (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.

Topping, K. & Ehly, S. (2001). Peer assisted learning: A framework for consultation. *Journal of Educational and Psychological Consultation, 12*(2), 113-132.

Tschannen-Moran, M. & Hoy, A. W. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education, 17*, 783-805.

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

VanLehn, K., Graesser, A., Jackson, G., Jordan, P., Olney, A., & Rose, C. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*, 3-62.

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

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Pres.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*, 663-676.